JC07 Rec'd PCT/PTO 27 APR 2001

(390 (000)) U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

ATTORNEY'S DOCKET NUMBER: Mo-6305/HR-199

U.S. APPLICATION NO (If known, see 37 CFR 1.5): 09/830514

INTERNATIONAL APPLICATION NO.	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/EP99/07952	October 20, 1999	October 31, 1998

TITLE OF INVENTION: Construction of Production Strains for Producing Substituted Phenols By Specifically Inactivating Genes of the Eugenol and Ferulic Acid Catabolism

APPLICANT(S) FOR DO/EO/US: RABENHORST, Jurgen; STEINBUCHEL, Alexander; PRIEFERT, Horst and OVERHAGE, Jorg

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. [X] This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. [ ] This is a **SECOND or SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. [X] This is an express request to begin national examination procedures (35 U.S.C. 371(f)). The submission must include items (5), (6), (9) and (21) indicated below.
4. [X] The US has been elected by the expiration of 19 months from the priority date (Article 31).
5. [X] A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. [X] is attached hereto (required only if not communicated by the International Bureau).
   b. [ ] has been communicated by the International Bureau.
   c. [ ] is not required, as the application was filed in the United States Receiving Office (RO/US).
6. [X] An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).
   a. [ ] is attached hereto.
   b. [ ] has been previously submitted under 35 U.S.C. 154(d)(4).
7. [ ] Amendments to the claims of the International Aplication under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. [ ] are attached hereto (required only if not communicated by the International Bureau).
   b. [ ] have been communicated by the International Bureau.
   c. [ ] have not been made; however, the time limit for making such amendments has NOT expired.
   d. [ ] have not been made and will not be made.
8. [ ] An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371 (c)(3)).
9. [X] An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. [ ] An English lanugage translation of the annexes of the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11 to 20 below concern document(s) or information included:

11. [ ] An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. [X] An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. [X] A FIRST preliminary amendment.
14. [ ] A SECOND or SUBSEQUENT preliminary amendment.
15. [ ] A substitute specification.
16. [ ] A change of power of attorney and/or address letter.
17. [X] A computer-readable form of the sequence listing in accordance with PCT Rule 13ter.2 and 35 U.S.C. 1.821 - 1.825.
18. [ ] A second copy of the published international application under 35 U.S.C. 154(d)(4).
19. [ ] A second copy of the English language translation of the international application under 35 U.S.C. 154(d)(4).
20. [X] Other items or information:

Preliminary Amendment w/Abstract, Sequence Listing (Paper Copy and Disk Copy)
Form PTO 1449 w/references

JC18 Rec'd PCT/PTO 27 APR 2001

U.S. APPLICATION NO (if known, see 37 CFR 1.5)	INTERNATIONAL APPLICATION NO	ATTORNEY'S DOCKET NUMBER
To Be Assigned 09/830514	PCT/EP99/07952	Mo-6305/HR-199

21. ☒ The following fees are submitted:

	CALCULATIONS PTO USE ONLY	
BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)):		
Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a (2)) paid to USPTO and International Search Report not prepared by the EPO or JPO **$1000.00**		
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO **$860.00**		
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO **$710.00**		
International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4) **$690.00**		
International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4) **$100.00**		
ENTER APPROPRIATE BASIC FEE AMOUNT =	$ 860.00	
Surcharge of **$130.00** for furnishing the oath or declaration later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(e)).	$ 0.00	

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE	$	
Total claims	15 -20 =	0	**x $18.00**	$ 0.00	
Independent claims	5 -3 =	2	**x $80.00**	$ 160.00	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			**+ $270.00**	$ 0.00	
TOTAL OF ABOVE CALCULATIONS =				$ 1,020.00	
☐ Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2. +				$ 0.00	
SUBTOTAL =				$ 1,020.00	
Processing fee of **$130.00** for furnishing the English translation later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(f)).				$ 0.00	
TOTAL NATIONAL FEE =				$ 1,020.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). **$40.00** per property +				$ 40.00	
TOTAL FEES ENCLOSED =				$ 1,060.00	
				Amount to be refunded:	$
				charged:	$

a. ☐ A check in the amount of $ _____ to cover the above fees is enclosed.

b. ☒ Please charge my Deposit Account No. 13-3848 in the amount of $ 1,060.00 to cover the above fees. A duplicate copy of this sheet is enclosed.

c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 13-3848. A duplicate copy of this sheet is enclosed.

d. ☐ Fees are to be charged to a credit card. **WARNING:** Information on this form may become public. **Credit card information should not be included on this form.** Provide credit card information and authorization on PTO-2038.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137 (a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO

00157
PATENT TRADEMARK OFFICE

SIGNATURE

Noland J. Cheung
NAME

39,138
REGISTRATION NUMBER

JC18 Rec'd PCT/PTO 27 APR 2001

TRANSMITTAL LETTER TO THE UNITED STATES RECEIVING OFFICE

Date	April 27, 2001
International Application No.	PCT/EP99/07952 09/830514
Attorney Docket No.	Mo-6305/HR-199

I. Certification under 37 CFR 1.10 (if applicable)

ET146893673US	April 27, 2001
Express Mail mailing number	Date of Deposit

I hereby certify that the application/correspondence attached hereto is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to Assistant Commissioner for Patents, Washington, D.C. 20231.

Donna J. Veatch	Donna J. Veatch
Signature of person mailing correspondence	Typed or printed name of person mailing correspondence

II. [X] New International Application

TITLE	CONSTRUCTION OF PRODUCTION STRAINS FOR PRODUCING SUBSTITUTED PHENOLS BY SPECIFICALLY INACTIVATING GENES OF THE EUGENOL AND FERULIC ACID CATABOLISM	Earliest priority date (Day/Mon/Year): (31/10/98)

SCREENING DISCLOSURE INFORMATION: In order to assist in screening the accompanying international application for purposes of determining whether a license for foreign transmittal should and could be granted and for other purposes. the following information is supplied. (Note: check as many boxes as apply):

A. [X] The invention disclosed was not made in the United States.

B. [X] There is no prior U.S. application relating to this invention.

C. [ ] The following prior U.S. application(s) contain subject matter which is related to the invention disclosed in the attached international application. *(NOTE. priority to these applications may or may not be claimed on form PCT/RO/101 (Request) and this listing does not constitute a claim for priority.)*

application no.		filed on	
application no.		filed on	

D. [ ] The present international application [ ] contains additional subject matter not found in the prior U.S. application(s) identified in paragraph C. above. The additional subject matter is found on pages [ ] and [ ] **DOES NOT ALTER** [ ] **MIGHT BE CONSIDERED TO ALTER** the general nature of the invention in a manner which would require the U.S. application to have been made available for inspection by the appropriate defense agencies under 35 U.S.C. 181 and 37 CFR 5.1. See 37 CFR 5.15

III. [ ] A Response to an Invitation from the RO/US. The following document(s) is(are) enclosed:

A. [ ] A Request for An Extension of Time to File a Response

B. [ ] A Power of Attorney (General or Regular)

C. [ ] Replacement pages:

pages		of the request (PCT/RO/101)	pages		of the figures
pages		of the description	pages		of the abstract
pages		of the claims			

D. [ ] Submission of Priority Documents

Priority document		Priority document	

E. [ ] Fees as specified on attached Fee Calculation sheet form PCT/RO/101 annex

IV. [ ] A Request for Rectification under PCT 91 [ ] A Petition [ ] A Sequence Listing Diskette

V. [X] Other (please specify): Preliminary Amendment w/Abstract, Sequence Listing (Paper and Disk Copy) Form PTO 1449 w/references Drawings (3 sheets)

The person signing this form is the:		
[ ]	Applicant	Noland J. Cheung
[X]	Attorney/Agent (Reg. No.) 39,138	Typed name of signer
[ ]	Common Representative	Signature

PATENT APPLICATION
Mo-6305
HR-199

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION OF	
JÜRGEN RABENHORST, ET AL.	PCT/EP99/07952
SERIAL NUMBER: TO BE ASSIGNED	
FILED: HEREWITH	
TITLE: CONSTRUCTION OF PRODUCTION STRAINS FOR PRODUCING SUBSTITUTED PHENOLS BY SPECIFICALLY INACTIVATING GENES OF THE EUGENOL AND FERULIC ACID CATABOLISM	

PRELIMINARY AMENDMENT

Assistant Commissioner for Patents
Washington, D.C. 20231
Sir:

Upon the granting of a Serial Number and Filing Date and prior to the examination of the subject application, kindly amend the Specification and Claims as follows:

"Express Mail" mailing label number EL146893673US
Date of Deposit April 27, 2001

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Assistant Commissioner of Patents and Trademarks, Washington, D.C. 20231

Donna J. Veatch
(Name of person mailing paper or fee)

(Signature of person mailing paper or fee)

IN THE SPECIFICATION:

Kindly replace the Title of the Invention with the following:

-- CONSTRUCTION OF PRODUCTION STRAINS FOR PRODUCING SUBSTITUTED PHENOLS BY SPECIFICALLY INACTIVATING GENES OF THE EUGENOL AND FERULIC ACID CATABOLISM --.

Kindly insert the following "ABSTRACT" page

-- The present invention relates to a transformed and/or mutagenated unicellular or multicellular organism which is characterized in that enzymes of the eugenol and/or ferulic acid catabolism are deactivated in such a manner that the intermediates coniferyl alcohol, coniferyl aldehyde, ferulic acid, vanillin and/or vanillinic acid are accumulated. --

On page 1, line 4, kindly insert the following:

-- FIELD OF THE INVENTION --.

On page 1, line 7, kindly insert the following:

--BACKGROUND OF THE INVENTION--.

On page 2, after line 9, kindly insert the following:

-- BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1a to 1r show gene structures for isolating organisms and mutants.

FIG. 2a: shows a nucleotide sequence of the *calA*ΩKm gene structure (SEQ ID NO: 1).

FIG. 2b: shows a nucleotide sequence of the *calA*ΩGm gene structure (SEQ ID NO: 2).

FIG. 2c: shows a nucleotide sequence of the $calA_{\Delta}$ gene structure (SEQ ID NO: 3).

FIG. 2d: shows a nucleotide sequence of the *calB*ΩKm gene structure (SEQ ID NO: 4).

FIG. 2e: shows a nucleotide sequence of the *calB*ΩGm gene structure (SEQ ID NO: 5).

FIG. 2f: shows a nucleotide sequence of the $calB_{\Delta}$ gene structure (SEQ ID NO: 6).

FIG. 2g: shows a nucleotide sequence of the *fcs*ΩKm gene structure (SEQ ID NO: 7).

FIG. 2h: shows a nucleotide sequence of the *fcs*ΩGm gene structure (SEQ ID NO: 8).

FIG. 2i: shows a nucleotide sequence of the fcs_{Δ} gene structure (SEQ ID NO: 9).

FIG. 2j: shows a nucleotide sequence of the *ech*ΩKm gene structure (SEQ ID NO: 10).

FIG. 2k: shows a nucleotide sequence of the *ech*ΩGm gene structure (SEQ ID NO: 11).

FIG. 2l: shows a nucleotide sequence of the ech_{Δ} gene structure (SEQ ID NO: 12).

FIG. 2m: shows a nucleotide sequence of the *vdh*ΩKm gene structure (SEQ ID NO: 13).

FIG. 2n: shows a nucleotide sequence of the *vdh*ΩGm gene structure (SEQ ID NO: 14).

FIG. 2o: shows a nucleotide sequence of the vdh_{Δ} gene structure (SEQ ID NO: 15).

FIG. 2p: shows a nucleotide sequence of the *aat*ΩKm gene structure (SEQ ID NO: 16).

FIG. 2q: shows a nucleotide sequence of the *aat*ΩGm gene structure (SEQ ID NO: 17).

FIG. 2r: shows a nucleotide sequence of the aat_{Δ} gene structure (SEQ ID NO: 18). --.

On page 2, line 10, kindly insert the following:

--SUMMARY OF THE INVENTION--.

On page 2, line 19, kindly insert the following:

--DETAILED DESCRIPTION OF THE INVENTION--.

IN THE CLAIMS:

Kindly cancel Claims 1 - 16.

Kindly add the following new claims:

-- 17. Transformed and/or mutagenized unicellular or multicellular organism comprising enzymes of eugenol and/or ferulic acid catabolism which are inactivated such that the intermediates coniferyl alcohol, coniferyl aldehyde, ferulic acid, vanillin and/or vanillic acid accumulate.

18. An organism according to Claim 17, wherein eugenol and/or ferulic acid catabolism is altered by inserting Ω elements, or introducing deletions, into corresponding genes.

19. Organism according to Claim 17, wherein one or more genes encoding the enzymes coniferyl alcohol dehydrogenases, coniferyl aldehyde dehydrogenases, feruloyl-CoA synthetases, enoyl-CoA hydratase-aldolases, beta-ketothiolases, vanillin dehydrogenases or vanillic acid demethylases is/are altered and/or inactivated.

20. An organism according to Claim 17, wherein said organism is unicellular.

21. An organism according to Claim 20, wherein said organism is selected from a group consisting of a microorganism, a plant or animal cell.

22. An organism according to Claim 17, wherein said organism is a bacterium.

23. An organism according to Claim 22, wherein said organism is of the *Pseudomonas* species.

24. Gene structures comprising nucleotide sequences which encode the enzymes coniferyl alcohol dehydrogenases, coniferyl aldehyde dehydrogenases, feruloyl-CoA synthetases, enoyl-CoA hydratase-aldolases, beta-ketothiolases, vanillin-dehydrogenases or vanillic acid demethylases, or two or more of these enzymes, and are altered and/or inactivated.

25. Gene structures having the sequences corresponding to SEQ ID NO:1 to SEQ ID NO: 18.

26. Vectors comprising at least one gene structure having the sequences corresponding to SEQ ID NO:1 to SEQ ID NO: 18.

27. A transformed organism according to Claim 17, wherein said organism comprises at least one vector comprising at least one gene structure having the sequences corresponding to SEQ ID NO:1 to SEQ ID NO: 18.

28. Organism according to Claim 17, wherein said organism comprises at least one gene structure having the sequences corresponding to SEQ ID NO:1 to SEQ ID NO: 18 which is integrated into the genome instead of the respective intact gene.

29. Process for the biotechnological preparation of alcohols, aldehydes and organic acids, comprising the step of adding an organism comprising enzymes of eugenol and/or ferulic acid catabolism which are inactivated such that the intermediates coniferyl alcohol, coniferyl aldehyde, ferulic acid, vanillin and/or vanillic acid accumulate.

30. Process for preparing an organism according to Claim 17, wherein the alteration eugenol and/or ferulic acid catabolism is achieved by microbiological culturing methods.

31. Process for preparing an organism according to Claim 29, wherein the alteration in eugenol and/or ferulic acid catabolism, and/or the inactivation of the corresponding genes, is achieved by means of recombinant DNA methods. --.

REMARKS

The Applicants respectfully request the Preliminary Amendment be entered as the amendment places the claims as well as the Specification in proper form.

New Claims 17 - 31 replace now cancelled Claims 1 - 16. Support for the new claims are found in the respective original cancelled claims. The Applicants respectfully submit that no new matter is added.

Additionally, the Applicants hereby submit a paper copy of the "Sequence Listing" as well as a copy of the "Sequence Listing" in computer readable form. The "Sequence Listing" has been amended to place it in proper form for U.S. filing. The Applicants also state that the information recorded in computer readable form is identical to the written sequence listing.

The attached page is captioned "VERSION WITH MARKINGS TO SHOW CHANGES MADE".

Respectfully submitted,

By ______________________
Noland J. Cheung
Attorney for Applicants
Reg. No. 39,138

Bayer Corporation
100 Bayer Road
Pittsburgh, Pennsylvania 15205-9741
(412) 777-8338
FACSIMILE PHONE NUMBER:
(412) 777-8363
s:\ksl\NJC1008

VERSION WITH MARKINGS TO SHOW CHANGES MADE

IN THE SPECIFICATION:

Kindly replace the Title of the Invention with the following:

-- CONSTRUCTION OF PRODUCTION STRAINS FOR PRODUCING SUBSTITUTED PHENOLS BY SPECIFICALLY INACTIVATING GENES OF THE EUGENOL AND FERULIC ACID CATABOLISM --.

Kindly insert the following "ABSTRACT" page

-- The present invention relates to a transformed and/or mutagenated unicellular or multicellular organism which is characterized in that enzymes of the eugenol and/or ferulic acid catabolism are deactivated in such a manner that the intermediates coniferyl alcohol, coniferyl aldehyde, ferulic acid, vanillin and/or vanillinic acid are accumulated. --

On page 1, line 4, kindly insert the following:

-- FIELD OF THE INVENTION --.

On page 1, line 7, kindly insert the following:

--BACKGROUND OF THE INVENTION--.

On page 2, after line 9, kindly insert the following:

-- BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1a to 1r show gene structures for isolating organisms and mutants.

FIG. 2a: shows a nucleotide sequence of the *calA*ΩKm gene structure (SEQ ID NO: 1).

FIG. 2b: shows a nucleotide sequence of the *calA*ΩGm gene structure (SEQ ID NO: 2).

FIG. 2c: shows a nucleotide sequence of the $calA_{\Delta}$ gene structure (SEQ ID NO: 3).

FIG. 2d: shows a nucleotide sequence of the *calB*ΩKm gene structure (SEQ ID NO: 4).

FIG. 2e: shows a nucleotide sequence of the *calB*ΩGm gene structure (SEQ ID NO: 5).

FIG. 2f: shows a nucleotide sequence of the $calB_{\Delta}$ gene structure (SEQ ID NO: 6).

FIG. 2g: shows a nucleotide sequence of the *fcs*ΩKm gene structure (SEQ ID NO: 7).

FIG. 2h: shows a nucleotide sequence of the *fcs*ΩGm gene structure (SEQ ID NO: 8).

FIG. 2i: shows a nucleotide sequence of the fcs_{Δ} gene structure (SEQ ID NO: 9).

FIG. 2j: shows a nucleotide sequence of the *ech*ΩKm gene structure (SEQ ID NO: 10).

FIG. 2k: shows a nucleotide sequence of the *ech*ΩGm gene structure (SEQ ID NO: 11).

FIG. 2l: shows a nucleotide sequence of the ech_{Δ} gene structure (SEQ ID NO: 12).

FIG. 2m: shows a nucleotide sequence of the *vdh*ΩKm gene structure (SEQ ID NO: 13).

FIG. 2n: shows a nucleotide sequence of the *vdh*ΩGm gene structure (SEQ ID NO: 14).

FIG. 2o: shows a nucleotide sequence of the vdh_{Δ} gene structure (SEQ ID NO: 15).

FIG. 2p: shows a nucleotide sequence of the *aat*ΩKm gene structure (SEQ ID NO: 16).

FIG. 2q: shows a nucleotide sequence of the *aat*ΩGm gene structure (SEQ ID NO: 17).

FIG. 2r: shows a nucleotide sequence of the aat_{Δ} gene structure (SEQ ID NO: 18). --.

On page 2, line 10, kindly insert the following:

--SUMMARY OF THE INVENTION--.

On page 2, line 19, kindly insert the following:

--DETAILED DESCRIPTION OF THE INVENTION--.

IN THE CLAIMS:

Kindly cancel Claims 1 - 16.

Kindly add the following new claims:

-- 17. Transformed and/or mutagenized unicellular or multicellular organism comprising enzymes of eugenol and/or ferulic acid catabolism which are inactivated such that the intermediates coniferyl alcohol, coniferyl aldehyde, ferulic acid, vanillin and/or vanillic acid accumulate.

18. An organism according to Claim 17, wherein eugenol and/or ferulic acid catabolism is altered by inserting Ω elements, or introducing deletions, into corresponding genes.

19. Organism according to Claim 17, wherein one or more genes encoding the enzymes coniferyl alcohol dehydrogenases, coniferyl aldehyde dehydrogenases, feruloyl-CoA synthetases, enoyl-CoA hydratase-aldolases, beta-ketothiolases, vanillin dehydrogenases or vanillic acid demethylases is/are altered and/or inactivated.

20. An organism according to Claim 17, wherein said organism is unicellular.

21. An organism according to Claim 20, wherein said organism is selected from a group consisting of a microorganism, a plant or animal cell.

22. An organism according to Claim 17, wherein said organism is a bacterium.

23. An organism according to Claim 22, wherein said organism is of the *Pseudomonas* species.

24. Gene structures comprising nucleotide sequences which encode the enzymes coniferyl alcohol dehydrogenases, coniferyl aldehyde dehydrogenases, feruloyl-CoA synthetases, enoyl-CoA hydratase-aldolases, beta-ketothiolases, vanillin-dehydrogenases or vanillic acid demethylases, or two or more of these enzymes, and are altered and/or inactivated.

25. Gene structures having the sequences corresponding to SEQ ID NO:1 to SEQ ID NO: 18.

26. Vectors comprising at least one gene structure having the sequences corresponding to SEQ ID NO:1 to SEQ ID NO: 18.

27. A transformed organism according to Claim 17, wherein said organism comprises at least one vector comprising at least one gene structure having the sequences corresponding to SEQ ID NO:1 to SEQ ID NO: 18.

28. Organism according to Claim 17, wherein said organism comprises at least one gene structure having the sequences corresponding to SEQ ID NO:1 to SEQ ID NO: 18 which is integrated into the genome instead of the respective intact gene.

29. Process for the biotechnological preparation of alcohols, aldehydes and organic acids, comprising the step of adding an organism comprising enzymes of eugenol and/or ferulic acid catabolism which are inactivated such that the

intermediates coniferyl alcohol, coniferyl aldehyde, ferulic acid, vanillin and/or vanillic acid accumulate.

30. Process for preparing an organism according to Claim 17, wherein the alteration eugenol and/or ferulic acid catabolism is achieved by microbiological culturing methods.

31. Process for preparing an organism according to Claim 29, wherein the alteration in eugenol and/or ferulic acid catabolism, and/or the inactivation of the corresponding genes, is achieved by means of recombinant DNA methods. --.

CONSTRUCTION OF PRODUCTION STRAINS FOR PRODUCING SUBSTITUTED PHENOLS BY SPECIFICALLY INACTIVATING GENES OF THE EUGENOL AND FERULIC ACID CATABOLISM

ABSTRACT OF THE DISCLOSURE

The present invention relates to a transformed and/or mutagenated unicellular or multicellular organism which is characterized in that enzymes of the eugenol and/or ferulic acid catabolism are deactivated in such a manner that the intermediates coniferyl alcohol, coniferyl aldehyde, ferulic acid, vanillin and/or vanillinic acid are accumulated.

Constructing production strains for the preparation of substituted phenols by specifically inactivating genes of eugenol and ferulic acid catabolism

The present invention relates to the construction of production strains and to a process for preparing substituted methoxyphenols, in particular vanillin.

DE-A 4 227 076 (process for preparing substituted methoxyphenols, and microorganism which is suitable for this purpose) describes the preparation of substituted methoxyphenols using a novel *Pseudomonas* sp.. The starting material in this context is eugenol and the products are ferulic acid, vanillic acid, coniferyl alcohol and coniferyl aldehyde.

An extensive review of the biotransformations which were possible using ferulic acid, which was written by Rosazza et al. (Biocatalytic transformation of ferulic acid: an abundant aromatic natural product; J. Ind. Microbiol. **15:**457-471), also appeared in 1995.

The genes and enzymes for synthesizing coniferyl alcohol, coniferyl aldehyde, ferulic acid, vanillic and vanillin acid from *Pseudomonas* sp. were described in EP-A 0 845 532.

The enzymes for converting *trans*-ferulic acid into *trans*-feruloyl-SCoA ester and subsequently into vanillin, and also the gene for cleaving the ester, were described by the Institute of Food Research, Norwich, GB, in WO 97/35999. In 1998, the content of the patent also appeared in the form of scientific publications (Gasson et al. 1998. Metabolism of ferulic acid to vanillin. J. Biol. Chem. **273:**4163-4170; Narbad and Gasson 1998. Metabolism of ferulic acid via vanillin using a novel CoA-dependent pathway in a newly isolated strain of *Pseudomonas fluorescens*. Microbiology **144:**1397 - 1405).

HR 199

DE-A 195 32 317 describes the use of *Amycolatopsis* sp. for obtaining vanillin from ferulic acid fermentatively in high yields.

The known processes suffer from the disadvantage that they either achieve only very low yields of vanillin or make use of relatively expensive starting compounds. While the last-mentioned process (DE-A 195 32 317) does achieve high yields, the use of *Pseudomonas* sp. HR199 and *Amycolatopsis* sp. HR167 for biotransforming eugenol into vanillin requires a fermentation which is carried out in two steps, consequently leading to substantial expense and consumption of time.

The object of the present invention is therefore to construct organisms which are able to convert the relatively inexpensive raw material eugenol into vanillin in a one-step process.

This object is achieved by means of constructing production strains of unicellular or multicellular organisms, which strains are characterized in that enzymes of eugenol and/or ferulic acid catabolism are inactivated such that the intermediates coniferyl alcohol, coniferyl aldehyde, ferulic acid, vanillin and/or vanillic acid accumulate.

The production strain may be unicellular or multicellular. Accordingly, the invention can relate to microorganisms, plants or animals. Furthermore, use can also be made of extracts which are obtained from the production strain. According to the invention, preference is given to using unicellular organisms. These latter organisms can be microorganisms or animal or plant cells. According to the invention, particular preference is given to using fungi and bacteria. The highest preference is given to bacterial species. Those bacteria which may in particular be used, after their eugenol and/or ferulic acid catabolism has/have been altered, are species of *Rhodococcus*, *Pseudomonas* und *Escherichia*.

In the simplest case, known, conventional microbiological methods can be used for isolating the organisms which may be employed in accordance with the invention.

Thus, the enzymic activity of the proteins involved in eugenol and/or ferulic acid catabolism can be altered by using enzyme inhibitors. Furthermore, the enzymic activity of the proteins involved in eugenol and/or ferulic acid catabolism can be altered by mutating the genes which encode these proteins. Such mutations can be generated in a random manner by means of classical methods, for example by using UV irradiation or mutation-inducing chemicals.

Recombinant DNA methods, such as deletions, insertions and/or nucleotide exchanges, are likewise suitable for isolating the novel organisms. Thus, the genes of the organisms can, for example, be inactivated using other DNA elements (Ω elements). Suitable vectors can likewise be used for replacing the intact genes with gene structures which are altered and/or inactivated. In this context, the genes which are to be inactivated, and the DNA elements which are employed for the inactivation, can be obtained by means of classical cloning techniques or by means of polymerase chain reactions (PCR).

For example, in one possible embodiment of the invention, eugenol catabolism and ferulic acid catabolism can be altered by inserting Ω elements, or introducing deletions, into appropriate genes. In this context, the abovementioned recombinant DNA methods can be used to inactivate the functions of the genes, which encode dehydrogenases, synthetases, hydratase-adolases, thiolases or demethylases, such that production of the relevant enzymes is blocked. Preferably, the genes are those which encode coniferyl alcohol dehydrogenases, coniferyl aldehyde dehydrogenases, feruloyl-CoA synthetases, enoyl-CoA hydratase-aldolases, beta-ketothiolases, vanillin dehdrogenases or vanillic acid demethylases. Very particular preference is given to genes which encode the amino acid sequences specified in EP-A 0845532 and/or nucleotide sequences which encode their allelic variations.

The invention accordingly also relates to gene structures for preparing transformed organisms and mutants.

Preference is given to employing gene structures in which the nucleotide sequences encoding dehydrogenases, synthetases, hydratase-aldolases, thiolases or demethylases are inactivated for isolating the organisms and mutants. Particular preference is given to gene structures in which the nucleotide sequences encoding coniferyl alcohol dehydrogenases, coniferyl aldehyde dehydrogenases, feruloyl-CoA synthetases, enoyl-CoA hydratase-aldolases, beta-ketothiolases, vanillin dehydrogenases or vanillic acid demethylases are inactivated. Very particular preference is given to gene structures which exhibit the structures given in Figures 1a to 1r having the nucleotide sequences which are depicted in Figures 2a to 2r and/or nucleotide sequences encoding their allelic variations. In this context, particular preference is given to nucleotide sequences 1 to 18.

The invention also encompasses the part sequences of these gene structures as well as functional equivalents. Functional equivalents are to be understood as meaning those derivatives of the DNA in which individual nucleobases have been exchanged (wobble exchanges) without the function being altered. Amino acids may also be exchanged at the protein level without this resulting in an alteration in function.

One or more DNA sequences can be inserted upstream and/or downstream of the gene structures. By cloning the gene structures, it is possible to obtain plasmids or vectors which are suitable for the transformation and/or transfection of an organism and/or for conjugative transfer into an organism.

The invention furthermore relates to plasmids and/or vectors for preparing the organisms and mutants which are transformed in accordance with the invention. These organisms and mutants consequently harbour the gene structures which have been described. The present invention accordingly also relates to organisms which harbour the said plasmids and/or vectors.

The nature of the plasmids and/or vectors depends on what they are being used for. In order, for example, to be able to replace the intact genes of eugenol and/or ferulic

acid catabolism in pseudomonads with the genes which have been inactivated with omega elements, there is a need for vectors which, on the one hand, can be transferred into pseudomonads (conjugatively transferable plasmids) but which, on the other hand, cannot be replicated in these organisms and are consequently unstable in pseudomonads (so-called suicide plasmids). DNA segments which are transferred into pseudomonads with the aid of such a plasmid system can only be retained if they become integrated by homologous recombination into the genome of the bacterial cell.

The described gene structures, vectors and plasmids may be used for preparing different transformed organisms or mutants. The said gene structures can be used for replacing intact nucleic acid sequences with altered and/or inactivated gene structures. In the cells, which can be obtained by transformation or transfection or conjugation, the intact gene is replaced, by homologous recombination, with the altered and/or inactivated gene structure, as a consequence of which the resulting cells now only possess the altered and/or inactivated gene structure in their genome. In this way, preferably genes can be altered and/or inactivated, in accordance with the invention, such that the relevant organisms are able to produce coniferyl alcohol, coniferyl aldehyde, ferulic acid, vanillin and/or vanillic acid.

Mutants of the strain *Pseudomonas* sp. HR199 (DSM 7063), which was described in detail in DE-A 4 227 076 and EP-A 0845532, are examples of production strains which have been constructed in this way in accordance with the invention, with the corresponding gene structures ensuing, inter alia, from Figures 1a to 1r, in combination with Figures 2a to 2r:

1. *Pseudomonas* sp. HR199*calA*ΩKm, which contains the ΩKm-inactivated *calA* gene in place of the intact *calA* gene encoding coniferyl alcohol dehydrogenase (Fig. 1a; Fig. 2a).

2. *Pseudomonas* sp. HR199*calA*ΩGm, which contains the ΩGm-inactivated *calA* gene in place of the intact *calA* gene encoding coniferyl alcohol dehydrogenase (Fig. 1b; Fig. 2b).
3. *Pseudomonas* sp. HR199*calA*Δ, which contains the deletion-inactivated *calA* gene in place of the intact *calA* gene encoding coniferyl alcohol dehydrogenase (Fig. 1c; Fig. 2c).
4. *Pseudomonas* sp. HR199*calB*ΩKm, which contains the ΩKm-inactivated *calB* gene in place of the intact *calB* gene encoding coniferyl aldehyde dehydrogenase (Fig. 1d; Fig. 2d)
5. *Pseudomonas* sp. HR199*calB*ΩGm, which contains the ΩGm-inactivated *calB* gene in place of the intact *calB* gene encoding coniferyl aldehyde dehydrogenase (Fig. 1e; Fig. 2e).
6. *Pseudomonas* sp. HR199*calB*Δ, which contains the deletion-inactivated *calB* gene in place of the intact *calB* gene encoding coniferyl aldehyde dehydrogenase (Fig.1f; Fig. 2f).
7. *Pseudomonas* sp. HR199*fcs*ΩKm, which contains the ΩKm-inactivated *fcs* gene in place of the intact *fcs* gene encoding feruloyl-CoA synthetase (Fig.1g; Fig. 2g).
8. *Pseudomonas* sp. HR199*fcs*ΩGm, which contains the ΩGm-inactivated *fcs* gene in place of the intact *fcs* gene encoding feruloyl-CoA synthetase (Fig.1h; Fig. 2h).
9. *Pseudomonas* sp. HR199*fcs*Δ, which contains the deletion-inactivated *fcs* gene in place of the intact *fcs* gene encoding feruloyl-CoA synthetase (Fig.1i; Fig. 2i).
10. *Pseudomonas* sp. HR199*ech*ΩKm, which contains the ΩKm-inactivated *ech* gene in place of the intact *ech* gene encoding enoyl-CoA hydratase-aldolase (Fig.1j; Fig. 2j).
11. *Pseudomonas* sp. HR199*ech*ΩGm, which contains the ΩGm-inactivated *ech* gene in place of the intact *ech* gene encoding enoyl-CoA hydratase-aldolase (Fig.1k; Fig. 2k).

12. *Pseudomonas* sp. HR199*ech*Δ, which contains the deletion-inactivated *ech* gene in place of the intact *ech* gene encoding enoyl-CoA hydratase-aldolase (Fig.1l; Fig. 2l).

13. *Pseudomonas* sp. HR199*aat*ΩKm, which contains the ΩKm-inactivated *aat* gene in place of the intact *aat* gene ecnoding beta-ketothiolase (Fig. 1m; Fig. 2m).

14. *Pseudomonas* sp. HR199*aat*ΩGm, which contains the ΩGm-inactivated *aat* gene in place of the intact *aat* gene encoding beta-ketothiolase (Fig.1n; Fig. 2n).

15. *Pseudomonas* sp. HR199*aat*Δ, which contains the deletion-inactivated *aat* gene in place of the intact *aat* gene encoding beta-ketothiolase (Fig.1o; 2o).

16. *Pseudomonas* sp. HR199*vdh*ΩKm, which contains the ΩKm-inactivated *vdh* gene in place of the intact *vdh* gene encoding vanillin dehydrogenase (Fig.1p; Fig. 2p).

17. *Pseudomonas* sp. HR199*vdh*ΩGm, which contains the ΩGm-inactivated *vdh* gene in place of the intact *vdh* gene encoding vanillin dehydrogenase (Fig.1q; Fig. 2q).

18. *Pseudomonas* sp. HR199*vdh*Δ, which contains the deletion-inactivated *vdh* gene in place of the intact *vdh* gene encoding vanillin dehydrogenase (Fig.1r; Fig. 2r).

19. *Pseudomonas* sp. HR199*vdhB*ΩKm, which contains the ΩKm-inactivated *vdhB* gene in place of the intact *vdhB* gene encoding vanillin dehydrogenase II.

20. *Pseudomonas* sp. HR199*vdhB*ΩGm, which contains the ΩGm-inactivated *vdhB* gene in place of the intact *vdhB* gene encoding vanillin dehydrogenase II.

21. *Pseudomonas* sp. HR199*vdhB*Δ, which contains the deletion-inactivated *vdhB* gene in place of the intact *vdhB* gene encoding vanillin dehydrogenase II.

22. *Pseudomonas* sp. HR199*adh*ΩKm, which contains the ΩKm-inactivated *adh* gene in place of the intact *adh* gene encoding alcohol dehydrogenase.

23. *Pseudomonas* sp. HR199*adh*ΩGm, which contains the ΩGm-inactivated *adh* gene in place of the intact *adh* gene encoding alcohol dehydrogenase.

24. *Pseudomonas* sp. HR199*adh*Δ which contains the deletion-inactivated *adh* gene in place of the intact *adh* gene encoding alcohol dehydrogenase.

25. *Pseudomonas* sp. HR199*vanA*ΩKm, which contains the ΩKm-inactivated *vanA* gene in place of the intact *vanA* gene encoding the α-subunit of vanillic acid demethylase.

26. *Pseudomonas* sp. HR199*vanA*ΩGm, which contains the ΩGm-inactivated *vanA* gene in place of the intact *vanA* gene encoding the α-subunit of vanillic acid demethylase.

27. *Pseudomonas* sp. HR199*vanA*Δ, which contains the deletion-inactivated *vanA* gene in place of the intact *vanA* gene encoding the α-subunit of vanillic acid demethylase.

28. *Pseudomonas* sp. HR199*vanB*ΩKm, which contains the ΩKm-inactivated *vanB* gene in place of the intact *vanB* gene encoding the β-subunit of vanillic acid demethylase.

29. *Pseudomonas* sp. HR199*vanB*ΩGm, which contains the ΩGm-inactivated *vanB* gene in place of the intact *vanB* gene encoding the β-subunit of vanillic acid demethylase.

30. *Pseudomonas* sp. HR199*vanB*Δ, which contains the deletion-inactivated *vanB* gene in place of the intact *vanB* gene encoding the β-subunit of vanillic acid demethylase.

The invention additionally relates to a process for the biotechnological preparation of organic compounds. In particular, this process can be used to prepare alcohols, aldehydes and organic acids. The latter are preferably coniferyl alcohol, coniferyl aldehyde, ferulic acid, vanillin and vanillic acid.

The above-described organisms are employed in the novel process. The organisms which are very particularly preferred include bacteria, in particular the *Pseudomonas*

species. Specifically, the abovementioned *Pseudomonas* species can preferably be employed for the following processes:

1. *Pseudomonas* sp. HR199*calA*ΩKm, *Pseudomonas* sp. HR199*calA*ΩGm and *Pseudomonas* sp. HR199*calA*Δ for preparing coniferyl alcohol from eugenol.

2. *Pseudomonas* sp. HR199*calB*ΩKm, *Pseudomonas* sp. HR199*calB*ΩGm and *Pseudomonas* sp. HR199*calB*Δ for preparing coniferyl aldehyde from eugenol or coniferyl alcohol.

3. *Pseudomonas* sp. HR199*fcs*ΩKm, *Pseudomonas* sp. HR199*fcs*ΩGm, *Pseudomonas* sp. HR199fcsΔ, Pseudomonas sp. HR199*ech*ΩKm, *Pseudomonas* sp. HR199*ech*ΩGm and *Pseudomonas* sp. HR199*ech*Δ for preparing ferulic acid from eugenol or coniferyl alcohol or coniferyl aldehyde.

4. *Pseudomonas* sp. HR199*vdh*ΩKm, *Pseudomonas* sp. HR199*vdh*ΩGm, *Pseudomonas* sp. HR199*vdhΔ*, *Pseudomonas* sp. HR199*vdh*ΩGm*vdhB*ΩKm, *Pseudomonas* sp. HR199*vdh*ΩKm*vdhB*ΩGm, *Pseudomonas* sp. HR199*vdh*Δ *vdhB*ΩGm and *Pseudomonas* sp. HR199*vdh*Δ*vdhB*ΩKm for preparing vanillin from eugenol or coniferyl alcohol or coniferyl aldehyde or ferulic acid.

5. *Pseudomonas* sp. HR199*vanA*ΩKm, *Pseudomonas* sp. HR199*vanA*ΩGm, *Pseudomonas* sp. HR199*vanA*Δ, *Pseudomonas* sp. HR199*vanB*ΩKm, *Pseudomonas* sp. HR199*vanB*ΩGm and *Pseudomonas* sp. HR199*vanB*Δ for preparing vanillic acid from eugenol or coniferyl alcohol or coniferyl aldehyde or ferulic acid or vanillin.

Eugenol is the preferred substrate. However, it is also possible to add further substrates or even to replace the eugenol with another substrate.

Suitable nutrient media for the organisms which are employed in accordance with the invention are synthetic, semisynthetic or complex culture media. These media may comprise carbon-containing and nitrogen-containing compounds, inorganic salts, where appropriate trace elements, and vitamins.

Carbon-containing compounds which may be suitable are carbohydrates, hydrocarbons or organic standard chemicals. Examples of compounds which may preferably be used are sugars, alcohols or sugar alcohols, organic acids or complex mixtures.

The sugar is preferably glucose. The organic acids which may preferably be employed are citric or acetic acid. Examples of the complex mixtures are malt extract, yeast extract, casein or casein hydrolysate.

Inorganic compounds are suitable nitrogen-containing substrates. Examples of these are nitrates and ammonium salts. Organic nitrogen sources can also be used. These sources include yeast extract, soya bean meal, casein, casein hydrolysate and corn steep liquor.

Examples of the inorganic salts which may be employed are sulphates, nitrates, chlorides, carbonates and phosphates. The metals which the said salts contain are preferably sodium, potassium, magnesium, manganese, calcium, zinc and iron.

The temperature for the culture is preferably in the range from 5 to 100°C. The range from 15 to 60°C is particularly preferred, with 22 to 37°C being most preferred.

The pH of the medium is preferably 2 to 12. The range from 4 to 8 is particularly preferred.

In principle, any bioreactor known to the skilled person can be employed for carrying out the novel process. Preferential consideration is given to any appliance which is

suitable for submerged processes. This means that vessels which do or do not possess a mechanical mixing device may be employed in accordance with the invention. Examples of the latter are shaking apparatuses, and bubble column reactors or loop reactors. The former preferably include all the known appliances which are fitted with stirrers of any design.

The novel process can be carried out continuously or batchwise. The fermentation time required for achieving a maximum quantity of product depends on the specific nature of the organism employed. However, in principle, the fermentation times are between 2 and 200 hours.

The invention is explained in more detail below while referring to examples:

Mutants of the eugenol-utilizing strain *Pseudomonas* sp. HR199 (DSM 7063) were generated in a targeted manner by specifically inactivating genes of eugenol catabolism by means of inserting omega elements or introducing deletions. The omega elements employed were DNA segments which encoded resistances to the antibiotics kanamycin (ΩKm) and gentamycin (ΩGm). These resistance genes were isolated from Tn*5* and the plasmid pBBR1MCS-5 using standard methods. The genes *calA*, *calB*, *fcs*, *ech*, *aat*, *vdh*, *adh*, *vdhB*, *vanA* and *vanB*, which encode coniferyl alcohol dehydrogenase, coniferyl aldehyde dehydrogenase, feruloyl-CoA synthetase, enoyl-CoA hydratase-aldolase, beta-ketothiolase, vanillin dehdrogenase, alcohol dehydrogenase, vanillin dehdrogenase II and vanillic acid demethylase, were isolated from genomic DNA of the strain *Pseudomonas* sp. HR199 using standard methods and cloned into pBluescript SK$^-$. By means of digesting with suitable restriction endonucleases, DNA segments were removed from these genes (deletion) or substituted with Ω elements (insertion), resulting in the respective gene being inactivated. The genes which had been mutated in this manner were recloned into conjugatively transferable vectors and subsequently introduced into the strain *Pseudomonas* sp. HR199. Suitable selection was used to obtain transconjugants which had replaced the respective functional wild-type gene with the newly

introduced inactivated gene. The insertion and deletion mutants which were obtained in this way now only possessed the respective inactivated gene. This procedure was used to obtain both mutants possessing only one defective gene and multiple mutants, in which several genes had been inactivated in this manner. These mutants were employed for biotransforming

a) eugenol into coniferyl alcohol, coniferyl aldehyde, ferulic acid, vanillin and/or vanillic acid;

b) coniferyl alcohol into coniferyl aldehyde, ferulic acid, vanillin and/or vanillic acid;

c) coniferyl aldehyde into ferulic acid, vanillin and/or vanillic acid;

d) ferulic acid into vanillin and/or vanillic acid, and

e) vanillin into vanillic acid.

Materials and Methods

Conditions for growing the bacteria.

Strains of *Escherichia coli* were propagated at 37°C in Luria-Bertani (LB) or M9 mineral medium (J. Sambrook, E. F. Fritsch and T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2nd Edition., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York). Strains of *Pseudomonas* sp. were propagated at 30°C in Nutrient Broth (NB, 0.8%, wt/vol) or in mineral medium (MM) (H. G. Schlegel, et al. 1961. Arch. Mikrobiol. **38:**209-222) or HR mineral medium (HR-MM) (J. Rabenhorst, 1996. Appl. Microbiol. Biotechnol. **46:**470-474.). Ferulic acid, vanillin, vanillic acid and protocatechuic acid were dissolved in dimethyl sulphoxide and added to the respective medium to give a final concentration of 0.1% (wt/vol). Eugenol was either added directly to the medium to give a final concentration of 0.1% (vol/vol) or applied to filter paper (circular filter 595, Schleicher & Schuell, Dassel, Germany) in the lids of MM agar plates. When transconjugants and mutants of *Pseudomonas* sp. were being propagated, tetracycline, kanamycin and gentamycin were employed in final concentrations of 25 μg/ml, 100 μg/ml and 7.5 μg/ml, respectively.

Qualitative and quantitative detection of metabolic intermediates in culture supernatants.

Culture supernatants were analysed by high pressure liquid chromatography (Knauer HPLC) either directly or after dilution with doubly distilled H_2O. The chromatography was carried out on Nucleosil 100 C18 (7 μm, 250 x 4 mm). 0.1% (vol/vol) formic acid and acetonitrile was used as the solvent. The course of the gradient employed for eluting the substances was as follows:

00:00 - 06:30 → 26% acetonitrile
06:30 - 08:00 → 100% acetonitrile
08:00 - 12:00 → 100% acetonitrile
12:00 - 13:00 → 26% acetonitrile
13:00 - 18:00 → 26% acetonitrile

Purification of vanillin dehydrogenase II.

The purification was carried out at 4°C.

Crude extract.

Pseudomonas sp. HR199 cells which had been propagated on eugenol were washed in 10 mM sodium phosphate buffer, pH 6.0, then resuspended in the same buffer and disrupted by being passed twice through a French press (Amicon, Silver Spring, Maryland, USA) at a pressure of 1000 psi. The cell homogenate was subjected to an ultracentrifugation (1 h, 100,000 x g, 4°C), resulting in the soluble fraction of crude extract being obtained as the supernatant.

Anion exchange chromatography on DEAE Sephacel.

The soluble fraction of the crude extract was dialysed overnight against 10 mM sodium phosphate buffer, pH 6.0. The dialysate was loaded onto a DEAE-Sephacel column (2.6 cm x 35 cm, bed volume[BV]: 186 ml) which had been equilibrated with 10 mM sodium phosphate buffer, pH 6.0, and which had a flow rate of 0.8 ml/min. The column was rinsed with two BV of 10 mM sodium phosphate buffer, pH 6.0. The vanillin dehydrogenase II (VDH II) was eluted with a linear salt gradient of from 0 to 400 mM NaCl in 10 mM sodium phosphate buffer, pH 6.0 (750 ml). 10 ml fractions were collected. Fractions having a high VDH II activity were combined to form the DEAE pool.

Determining the vanillin dehydrogenase activity.

The VDH activity was determined at 30°C using an optical enzymic test. The reaction mixture, whose volume was 1 ml, contained 0.1 mmol of potassium phosphate (pH 7.1), 0.125 μmol of vanillin, 0.5 μmol of NAD, 1.2 μmol of pyruvate (Na salt), lactate dehydrogenase (1 U; from pig heart) and enzyme solution. The oxidation of vanillin was monitored at λ = 340 nm ($\varepsilon_{vanillin}$ = 11.6 $cm^2/\mu mol$). The enzyme activity was given in units (U), with 1 U corresponding to the quantity of enzyme which converts 1 μmol of vanillin per minute. The protein concentrations in

the samples were determined using the method of Lowry et al. (O. H. Lowry, N. J. Rosebrough, A. L. Farr and R. J. Randall. 1951. J. Biol. Chem. **193:**265-275).

Determining the coniferyl alcohol dehydrogenase activity.

The CADH activity was determined at 30°C using an optical enzymic test in accordance with Jaeger et al. (E. L. Jaeger, Eggeling and H. Sahm. 1981. Current Microbiology. **6:**333-336). The reaction mixture, whose volume was 1 ml, contained 0.2 mmol of tris/HCl (pH 9.0), 0.4 μmol of coniferyl alcohol, 2 μmol of NAD, 0.1 mmol of semicarbazide and enzyme solution. The reduction of NAD was monitored at λ = 340 nm (ε = 6.3 $cm^2/\mu mol$). The enzyme activity was given units (U), with 1 U corresponding to the quantity of enzyme which converts 1 μmol of substrate per minute. The protein concentrations in the samples were determined by the method of Lowry et al. (O. H. Lowry, N. J. Rosebrough, A. L. Farr and R. J. Randall. 1951. J. Biol. Chem. **193:**265-275).

Determining the coniferyl aldehyde dehydrogenase activity.

The CALDH activity was determined at 30°C using an optical enzymic test. The reaction mixture, whose volume was 1 ml, contained 0.1 mmol of tris/HCl (pH 8.8), 0.08 μmol of coniferyl aldehyde, 2.7 μmol of NAD and enzyme solution. The oxidation of coniferyl aldehyde to ferulic acid was monitored at λ = 400 nm (ε = 34 $cm^2/\mu mol$). The enzymic activity was given in units (U) with 1 U corresponding to the quantity of enzyme which converts 1 μmol of substrate per minute. The protein concentrations in the samples were determined by the method of Lowry et al. (O. H. Lowry, N. J. Rosebrough, A. L. Farr and R. J. Randall. 1951. J. Biol. Chem. **193:**265-275).

Determining the feruloyl-CoA synthetase (ferulic acid thiokinase) activity.

The FCS activity was determined at 30°C using an optical enzymic test which was a modification of that of Zenk et al. (Zenk et al. 1980. Anal. Biochem. **101:**182-187). The reaction mixture, whose volume was 1 ml, contained 0.09 mmol of potassium phosphate (pH 7.0), 2.1 μmol of $MgCl_2$, 0.7 μmol of ferulic acid, 2 μmol of ATP,

0.4 μmol of coenzyme A and enzyme solution. The formation of the CoA ester from ferulic acid was monitored at λ = 345 nm (ε = 10 $cm^2/\mu mol$). The enzymic activity was given in units (U), with 1 U corresponding to the quantity of enzyme which converts 1 μmol of substrate per minute. The protein concentrations in the samples were determined using the method of Lowry et al. (O. H. Lowry, N. J. Rosebrough, A. L. Farr and R. J. Randall. 1951. J. Biol. Chem. **193:**265-275).

Electrophoretic methods.

Protein-containing extracts were fractionated under native conditions in 7.4% (wt/vol) polyacrylamide gels using the method of Stegemann et al. (Stegemann et al. 1973. Z. Naturforsch. **28c:**722-732) and under denaturing conditions in 11.5% (wt/vol) polyacrylamide gels using the method of Laemmli (Laemmli, U. K. 1970. Nature (London) **227:**680-685). Serva Blue R was used for non-specific protein staining. For specifically staining the coniferyl alcohol dehydrogenase, coniferyl aldehyde dehydrogenase and vanillin dehydrogenase, the gels were rebuffered for 20 min in 100 mM KP buffer (pH 7.0) and subequently incubated at 30°C in the same buffer to which 0.08% (wt/vol) NAD, 0.04% (wt/vol) p-nitro blue tetrazolium chloride, 0.003% (wt/vol) phenazine methosulphate and 1 mM of the respective substrate had been added until corresponding colour bands became visible.

Transfer of proteins from polyacrylamide gels to PVDF membranes.

Proteins were transferred from SDS-polyacrylamide gels to PVDF membranes (Waters-Millipore, Bedford, Mass., USA) using a Semidry Fastblot appliance (B32/33, Biometra, Göttingen, Germany) in accordance with the manufacturer's instructions.

Determining N-terminal amino acid sequences.

N-terminal amino acid sequences were determined using a Protein Peptide Sequencer (Type 477 A, Applied Biosystems, Foster City, USA) and a PTH analyser in accordance with the manufacturer's instructions.

Isolating and manipulating DNA

Genomic DNA was isolated using the method of Marmur (J. Marmur, 1961. J. Mol. Biol. **3:**208-218). Other plasmid DNA and/or DNA restriction fragments was/were isolated and analysed using standard methods (J. E. Sambrook, F. Fritsch and T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2nd Edition., Cold Spring Harbor Laboratory Press, Cold Spring Habor, New York).

Transferring DNA.

Competent *Escherichia coli* cells were prepared and transformed using the method of Hanahan (D. Hanahan, 1983. J. Mol. Biol. **166:**557-580). Conjugative plasmid transfer between plasmid-harbouring *Escherichia coli* S17-1 strains (donor) and *Pseudomonas* sp.strains (recipient) was performed on NB agar plates in accordance with the method of Friedrich et al. (B. Friedrich et al. 1981. J. Bacteriol. **147:**198-205), or by means of a "minicomplementation method" on MM agar plates containing 0.5% (wt/vol) gluconate as the C source and 25 μg of tetracycline/ml or 100 μg of kanamycin/ml. In this case, cells of the recipient were applied in one direction as an inoculation streak. After 5 min, cells of the donor strains were then applied as inoculation streaks, with these streaks crossing the recipient inoculation streak. After incubating at 30°C for 48 h, the transconjugants grew directly downstream of the crossing site whereas neither the donor strain nor the recipient strain was able to grow.

Hybridization experiments.

DNA restriction fragments were fractionated electrophoretically in a 0.8% (wt/vol) agarose gel in 50 mM tris- 50 mM boric acid- 1.25 mM EDTA buffer (pH 8.5) (J. E. Sambrook, F. Fritsch and T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.). The transfer of the denatured DNA out of the gel onto a positively charged nylon membrane (pore size: 0.45 μm, Pall Filtrationstechnik, Dreieich, Germany), the subsequent hybridization with biotinylated or digoxigenin-labelled DNA probes, and the preparation of these DNA probes, were all performed using standard methods

(J. E. Sambrook, F. Fritsch and T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York).

DNA sequencing.

Nucleotide sequences were determined "non-radioactively" in accordance with the Sanger et al. (Sanger et al. 1977. Proc. Natl. Acad. Sci. USA **74:**5463-5467) dideoxy chain termination method using a "LI-COR" DNA Sequencer Model 4000L" (LI-COR Inc., Biotechnology Division, Lincoln, NE, USA) and using a "thermo sequenase fluorescent labelled primer cycle sequencing kit with 7-deaza-dGTP" (Amersham Life Science, Amersham International plc., Little Chalfont, Buckinghamshire, England), in each case in accordance with the manufacturer's instructions.

Synthetic oligonucleotides were used to carry out sequencing in accordance with the "primer-hopping strategy" of Strauss et al. (E. C. Strauss et al. 1986. Anal. Biochem. **154:**353-360).

Chemicals, biochemicals and enzymes.

Restriction enzymes, T4 DNA ligase, lambda DNA and enzymes and substrates for the optical enzymic tests were obtained from C.F. Boehringer & Söhne (Mannheim, Germany) or from GIBCO/BRL (Eggenstein, Germany). [γ-^{32}P]ATP was from Amersham/Buchler (Braunschweig, Germany). Oligonucleotides were obtained from MWG-Biotech GmbH (Ebersberg, Germany). Type NA agarose was obtained from Pharmacia-LKB (Uppsala, Sweden). All other chemicals were from Haarmann & Reimer (Holzminden, Germany), E. Merck AG (Darmstadt, Germany), Fluka Chemie (Buchs, Switzerland), Serva Feinbiochemica (Heidelberg, Germany) or Sigma Chemie (Deisenhofen, Germany).

Examples

Example 1

Constructing omega elements which mediate resistances to kanamycin (Ω Km) or gentamycin (ΩGm).

For constructing the ΩKm element, the 2099 bp *Bgl*I fragment of Transposons Tn*5* (E. A. Auerswald, G. Ludwig and H. Schaller. 1981. Cold Spring Harb. Symp. Quant. Biol. **45:**107-113; E. Beck, G. Ludwig, E. A. Auerswald, B. Reiss and H. Schaller. 1982. Genes **19:**327-336; P. Mazodier, P. Cossart, E. Giraud and F. Gasser. 1985. Nucleic Acids Res. **13:**195-205) was isolated on a preparative scale. The fragment was shortened down to approx. 990 bp by treating it with Bal 31 nuclease. This fragment, which now only comprised the kanamycin resistance gene (encoding an aminoglycoside-3'-O-phosphotransferase), was then ligated to *Sma*I-cut pSKsym DNA (pBluescript SK^- derivative which contains a symmetrically constructed multiple cloning site [*Sal*I, *Hin*dIII, *Eco*RI, *Sma*I, *Eco*RI, *Hin*dIII, *Sal*I]). It was possible to reisolate the ΩKm element from the resulting plasmid as a *Sma*I fragment, an *Eco*RI fragment, a *Hin*dIII fragment or a *Sal*I fragment.

For constructing the ΩGm element, the 983 bp *Eae*I fragment of the plasmid pBR1MCS-5 (M. E. Kovach, P. H. Elzer, D. S. Hill, G. T. Robertson, M. A. Farris, R. M. Roop and K. M. Peterson. 1995. Genes **166:**175-176) was isolated on a preparative scale and then treated with mung bean nuclease (progressive digestion of single-stranded DNA molecule ends). This fragment, which now only comprised the gentamycin resistance gene (encoding a gentamycin-3-acetyltransferase), was then ligated to *Sma*I-cleaved pSKsym DNA (see above). It was possible to reisolate the ΩGm element from the resulting plasmid as a *Sma*I fragment, an *Eco*RI fragment, a *Hin*dIII fragment or a *Sal*I fragment.

Example 2

Cloning the genes from *Pseudomonas* sp. HR199 (DSM7063) which were to be inactivated by inserting Ω elements or by means of deletions.

The *fcs*, *ech*, *vdh* and *aat* genes were cloned separately proceeding from the *E.* coli S17-1 strains DSM 10439 and DSM 10440 and using the plasmids pE207 and pE5-1 (see EP-A 0845532). The given fragments were isolated on a preparative scale from these plasmids and treated as described below:

For cloning the *fcs* gene, the 2350 bp *SalI*/*Eco*RI fragment from plasmid pE207 and the 3700 bp *Eco*RI/*Sal*I fragment from plasmid pE5-1 were cloned together in pBluescript SK$^-$ such that the two fragments were joined together by way of the *Eco*RI ends. The 6050 bp *Sal*I fragment was isolated on a preparative scale from the resulting hybrid plasmid and shortened down to approx. 2480 bp by being treated with Bal 31 nuclease. *Pst*I linkers were subsequently ligated to the ends of the fragment and, after digestion with *Pst*I, the fragment was cloned into pBluescript SK$^-$ (pSK*fcs*). After transformation of *E. coli* XL1 blue, clones were obtained which expressed the *fcs* gene and exhibited an FCS activity of 0.2 U/mg of protein.

For cloning the *ech* gene, the 3800 bp *Hin*dIII/*Eco*RI fragment from plasmid pE207 was isolated on a preparative scale and shortened down to approx. 1470 bp by treating it with Bal 31 nuclease. *Eco*RI linkers were then ligated to the ends of the fragment and, after digestion with *Eco*RI, the fragment was cloned into pBluescript SK$^-$ (pSK*ech*).

For cloning the *vdh* gene, the 2350 bp *SalI*/*Eco*RI fragment from plasmid pE207 was isolated on a preparative scale. After cloning into pBluescript SK$^-$, the fragment was truncated at one end by approx. 1530 bp using an exonuclease III/mung bean nuclease system. An *Eco*RI linker was then ligated to the end of the fragment and, after digestion with *Eco*RI, the fragment was cloned into pBluescript SK$^-$ (pSK*vdh*).

Following transformation of *E. coli* XL1 blue, clones were obtained which expressed the VDH gene and exhibited a VDH activity of 0.01 U/mg of protein.

For cloning the *aat* gene, the 3700 bp *Eco*RI/*Sal*I fragment from plasmid pE5-1 was isolated on a preparative scale and shortened down to approx. 1590 bp by treating it with Bal 31 nuclease. *Eco*RI linkers were then ligated to the ends of the fragment and, after digestion with *Eco*RI, the fragment was cloned into pBluescript SK$^-$ (pSK*aat*).

Example 3

Inactivating the above-described genes by inserting Ω elements or by deleting constituent regions of these genes.

Plasmid pSK*fcs*, which contained the *fcs* gene, was digested with *Bss*HII, resulting in a 1290 bp fragment being excised from the *fcs* gene. Following religation, the deletion derivative of the *fcs* gene (*fcs*Δ) (see Figs. 1i and 2i) was obtained in cloned form in pBluescript SK$^-$ (pSK*fcs*Δ). In addition, after the fragment had been excised, the omega elements ΩKm and ΩGm were ligated in in its stead. This resulted in the Ω-inactivated derivatives of the *fcs* gene (*fcs*ΩKm, see Figs. 1g and 2g) and (*fcs*ΩGm, see Fig. 1h and 2h) being obtained in cloned form in pBluescript SK$^-$ (pSK*fcs*ΩKm and pSK*fcs*ΩGm). It was not possible to detect any FCS activity in crude extracts of the resulting *E. coli* clones, whose hybrid plasmids possessed an *fcs* gene which was inactivated by deletion or by Ω element insertion.

Plasmid pSK*ech*, which contained the *ech* gene, was digested with *Nru*I, resulting in a 53 bp fragment and a 430 bp fragment being excised from the *ech* gene. After religation, the deletion derivative of the *ech* gene (*ech*Δ, see Fig. 1l and 2l) was obtained in cloned form in pBluescript SK$^-$ (pSK*ech*Δ). In addition, after the fragments had been excised, the omega elements ΩKm and ΩGm were ligated in in their stead. This resulted in the Ω-inactivated derivatives of the *ech* gene (*ech*ΩKm

and *ech*ΩGm) being obtained in cloned form in pBluescript SK$^-$ (pSK*ech*ΩKm and pSK*ech*ΩGm).

Plasmid pSK*vdh*, which contained the *vdh* gene, was digested with *Bss*HII, resulting in a 210 bp fragment being excised from the *vdh* gene. After religation, the deletion derivative of the *vdh* gene (*vdh*Δ, see Figs. 1o and 2o) was obtained in cloned form in pBluescript SK$^-$ (pSK*vdh*Δ). In addition, after the fragment had been excised, the omega elements ΩKm and ΩGm were ligated in in its stead. This resulted in the Ω-inactivated derivatives of the *vdh* gene (*vdh*ΩKm and *vdh*ΩGm) being obtained in cloned form in pBluescript SK$^-$ (pSK*vdh*ΩKm, see Figs. 1m and 2m) and (pSK*vdh*ΩGm, see Figs. 1n and 2n). It was not possible to detect any VDH activity in crude extracts of the resulting *E. coli* clones, whose hybrid plasmids possessed a *vdh* gene which was inactivated by deletion or by Ω element insertion.

Plasmid pSK*aat*, which contained the *aat* gene, was digested with *Bss*HII, resulting in a 59 bp fragment being excised from the *aat* gene. After religation, the deletion derivative of the *aat* gene (*aat*Δ, see Figs. 1r and 2r) was obtained in cloned form in pBluescript SK$^-$ (pSK*aat*Δ). In addition, after the fragment had been excised, the omega elements ΩKm and ΩGm were ligated in in its stead. This resulted in the Ω-inactivated derivatives of the *aat* gene (*aat*ΩKm, see Figs. 1p and 2p) and (*aat*ΩGm, see Figs. 1q and 2q) being obtained in cloned form in pBluescript SK$^-$ (pSK*aat*ΩKm and pSK*aat*ΩGm).

Example 4

Subcloning the Ω element-inactivated genes into the conjugatively transferable "suicide plasmid" pSUP202.

In order to be able to replace the intact genes in *Pseudomonas sp.* HR199 with the Ω-element inactivated genes, there is a need for a vector which can, on the one hand, be transferred into pseudomonads (conjugatively transferable plasmids) but which, on the other hand, cannot replicate in these bacteria and is consequently unstable in pseudomonads ("suicide plasmid"). DNA segments which are transferred into pseudomonads using such a plasmid system can only be retained if they are integrated by means of homologous recombination (RecA-dependent recombination) into the genome of the bacterial cell. In the present case, the "suicide plasmid" pSUP202 (Simon et al. 1983. *In*: A. Pühler. Molecular genetics of the bacteria-plant interaction. Springer Verlag, Berlin, Heidelberg, New York, pp. 98-106) was used.

Following digestion with *Pst*I, the inactivated genes *fcs*ΩKm and *fcs*ΩGm were isolated from plasmids pSK*fcs*ΩKm and pSK*fcs*ΩGm and ligated to *Pst*I-cleaved pSUP202 DNA. The ligation mixtures were transformed into *E. coli* S17-1. Selection took place on tetracycline-containing LB medium which also contained kanamycin or gentamycin, respectively. Kanamycin-resistant transformants whose hybrid plasmid (pSUP*fcs*ΩKm) contained the inactivated gene *fcs*ΩKm were obtained. The corresponding hybrid plasmid (pSUP*fcs*ΩGm) of the gentamycin-resistant transformants contained the inactivated gene *fcs*ΩGm.

Following *Eco*RI digestion, the inactivated genes echΩKm and *ech*ΩGm were isolated from plasmids pSK*ech*ΩKm and pSK*ech*ΩGm and ligated to *Eco*RI-cleaved pSUP202 DNA. The ligation mixtures were transformed into *E. coli* S17-1. Selection took place on tetracycline-containing LB medium which also contained kanamycin or gentamycin, respectively. Kanamycin-resistant transformants whose hybrid plasmid (pSUP*ech*ΩKm) contained the inactivated gene *ech*ΩKm were obtained. The

corresponding hybrid plasmid (pSUP*ech*ΩGm) of the gentamycin-resistant transformants contained the inactivated gene *ech*ΩGm.

Following *Eco*RI digestion, the inactivated genes *vdh*ΩKm and *vdh*ΩGm were isolated from plasmids pSK*vdh*ΩKm and pSK*vdh*ΩGm and ligated to *Eco*RI-cleaved pSUP202 DNA. The ligation mixtures were transformed into *E. coli* S17-1. Selection took place on tetracycline-containing LB medium which also contained kanamycin or gentamycin, respectively. Kanamycin-resistant transformants whose hybrid plasmid (pSUP*vdh*ΩKm) contained the inactivated gene *vdh*ΩKm were obtained. The corresponding hybrid plasmid (pSUP*vdh*ΩGm) of the gentamycin-resistant transformants contained the inactivated gene *vdh*ΩGm.

Following *Eco*RI digestion, the inactivated genes *aat*ΩKm and *aat*ΩGm were isolated from plasmids pSK*aat*ΩKm and pSK*aat*ΩGm and ligated to *Eco*RI-cleaved pSUP202 DNA. The ligation mixtures were transformed into *E. coli* S17-1. Selection took place on tetracycline-containing LB medium which also contained kanamycin or gentamycin, respectively. Kanamycin-resistant transformants whose hybrid plasmid (pSUP*aat*ΩKm) contained the inactivated gene *aat*ΩKm were obtained. The corresponding hybrid plasmid (pSUP*aat*ΩGm) of the gentamycin-resistant transformants contained the inactivated gene *aat*ΩGm.

Example 5

Subcloning the deletion-inactivated genes into the conjugatively transferable "suicide plasmid" PHE55, which possesses the "*sacB* selection system".

In order to be able to replace the intact genes in *Pseudomonas* sp. HR199 with the deletion-inactivated genes, there is a need for a vector which possesses the properties which have already been described in the case of pSUP202. Since no possibility (no antibiotic resistance) exists of selecting for successful replacement of the genes in *Pseudomonas* sp. HR199 in the case of deletion-inactivated genes, in contrast to the Ω element-inactivated genes, another selection system had to be used. In the "*sacB*

selection system", the replacing, deletion-inactivated gene is cloned in a plasmid which possesses the *sacB* gene in addition to an antibiotic resistance gene. Following the conjugative transfer of this hybrid plasmid into a pseudomonad, the plasmid is integrated by means of homologous recombination at the site in the genome at which the intact gene is located (first crossover). This results in a "heterogenotic" strain which possesses both an intact gene and a deletion-inactivated gene, with these genes being separated from each other by the pHE55 DNA. These strains exhibit the resistance which is encoded by the vector and also possess an active *sacB* gene. The intention then is that the pHE55 DNA, together with the intact gene, should then be separated out of the genomic DNA by means of a second homologous recombination event (second crossover). This recombination event results in a strain which now only possesses the inactivated gene. In addition, the pHE55-coded antibiotic resistance and the *sacB* gene are both lost. If strains are streaked on sucrose-containing media, the growth of strains which express the *sacB* gene is inhibited since the gene product converts sucrose into a polymer which is accumulated in the periplasm of the cells. The growth of cells which no longer carry the *sacB* gene as a result of the second recombination event having taken place is consequently not inhibited. In order to have a possibility of selecting phenotypically for the integration of the deletion-inactivated gene, this gene is not exchanged for an intact gene; instead, use is made of a strain in which the gene to be replaced is already "labelled" by the insertion of an Ω element. When successful replacement takes place, the resulting strain loses the antibiotic resistance which is encoded by the Ω element.

Following digestion with *Pst*I, the inactivated gene *fcs*Δ was isolated from plasmid pSK*fcs*Δ and ligated to *Pst*I-cleaved pHE55 DNA. The ligation mixture was transformed into *E. coli* S17-1. Selection took place on tetracycline-containing LB medium. Tetracycline-resistant transformants, whose hybrid plasmid (pHE*fcs*Δ) contained the inactivated gene *fcs*Δ, were obtained.

Following digestion with *Eco*RI, the inactivated gene *ech*Δ was isolated from plasmid pSK*ech*Δ and treated with mung bean nuclease (generation of blunt ends).

The fragment was ligated to *Bam*HI-cleaved and mung bean nuclease-treated pHE55 DNA. The ligation mixture was transformed into *E. coli* S17-1. Selection took place on tetracycline-containing LB medium. Tetracycline-resistant transformants, whose hybrid plasmid (pHE*ech*Δ) contained the inactivated gene *ech*Δ, were obtained

Following digestion with *Eco*RI, the inactivated gene *vdh*Δ was isolated from plasmid pSK*vdh*Δ and treated with mung bean nuclease. The fragment was ligated to *Bam*HI-cleaved and mung bean nuclease-treated pHE55 DNA. The ligation mixture was transformed into *E. coli* S17-1. Selection took place on tetracycline-containing LB medium. Tetracycline-resistant transformants, whose hybrid plasmid (pHE*vdh*Δ) contained the inactivated gene *vdh*Δ, were obtained.

Following digestion with *Eco*RI, the inactivated gene *aat*Δ was isolated from plasmid pSK*aat*Δ and treated with mung bean nuclease. The fragment was ligated to *Bam*HI-cleaved and mung bean nuclease-treated pHE55 DNA. The ligation mixture was transformed into *E. coli* S17-1. Selection took place on tetracycline-containing LB medium. Tetracycline-resistant transformants, whose hybrid plasmid (pHE*aat*Δ) contained the inactivated gene *aat*Δ, were obtained.

Example 6

Generating mutants of the strain *Pseudomonas* sp. HR199 in which genes of eugenol catabolism have been specifically inactivated by inserting an Ω-element.

The strain *Pseudomonas* sp. HR199 was employed as the recipient in conjugation experiments in which strains of *E. coli* S17-1 harbouring the hybrid plasmids of pSUP202 which are listed below were used as donors. The transconjugants were selected on gluconate-containing mineral medium which contained the antibiotic corresponding to the Ω element. It was possible to distinguish between "homogenotic" (replacement of the intact gene with the Ω element insertion-inactivated gene by means of a double crossover) and "heterogenotic" (integration of the hybrid plasmid into the genome by means of a single crossover) transconjugants on the basis of the pSUP202-encoded tetracycline resistance.

The mutants *Pseudomonas* sp. HR199 *fcs*ΩKm and *Pseudomonas* sp. HR199 *fcs*ΩGm were obtained after conjugating *Pseudomonas* sp. HR199 with *E. coli* S17-1 (pSUP*fcs*ΩKm) and *E. coli* S17-1 (pSUP*fcs*ΩGm), respectively. The replacement of the intact *fcs* gene with the ΩKm-inactivated or ΩGm-inactivated gene (*fcs*ΩKm and *fcs*ΩGm, respectively) was verified by means of DNA sequencing.

The mutants *Pseudomonas* sp. HR199 *ech*ΩKm and *Pseudomonas* sp. HR199 *ech*ΩGm were obtained after conjugating *Pseudomonas* sp. HR199 with *E. coli* S17-1 (pSUP*ech*ΩKm) and *E. coli* S17-1 (pSUP*ech*ΩGm), respectively. The replacement of the intact *ech* gene with the ΩKm-inactivated or ΩGm-inactivated gene (*ech*ΩKm and *ech*ΩGm, respectively) was verified by means of DNA sequencing.

The mutants *Pseudomonas* sp. HR199 *vdh*ΩKm and *Pseudomonas* sp. HR199 *vdh*ΩGm were obtained after conjugating *Pseudomonas* sp. HR199 with *E. coli* S17-1 (pSUP*vdh*ΩKm) and *E. coli* S17-1 (pSUP*vdh*ΩGm), respectively. The

replacement of the intact *vdh* gene with the ΩKm-inactivated or ΩGm-inactivated gene (*vdh*ΩKm and *vdh*ΩGm, respectively) was verified by means of DNA sequencing.

The mutants *Pseudomonas* sp. HR199 *aat*ΩKm and *Pseudomonas* sp. HR199 *aat*ΩGm were obtained after conjugating *Pseudomonas* sp. HR199 with *E. coli* S17-1 (pSUP*aat*ΩKm) and *E. coli* S17-1 (pSUP*aat*ΩGm), respectively. The replacement of the intact *aat* gene with the ΩKm-inactivated or ΩGm-inactivated gene (*aat*ΩKm and *aat*ΩGm, respectively) was verified by means of DNA sequencing.

The mutant *Pseudomonas* sp. HR199 *fcs*ΩKm*vdh*ΩGm was obtained after conjugating *Pseudomonas* sp. HR199 *fcs*ΩKm with *E. coli* S17-1 (pSUP*vdh*ΩGm). The replacement of the intact *vdh* gene with the ΩGm-inactivated gene (*vdh*ΩGm) was verified by means of DNA sequencing.

The mutant *Pseudomonas* sp. HR199 *vdh*ΩKm*aat*ΩGm was obtained after conjugating *Pseudomonas* sp. HR199 *vdh*ΩKm with *E. coli* S17-1 (pSUP*aat*ΩGm). The replacement of the intact *aat* gene with the ΩGm-inactivated gene (*aat*ΩGm) was verified by means of DNA sequencing.

The mutant *Pseudomonas* sp. HR199 *vdh*ΩKm*ech*ΩGm was obtained after conjugating *Pseudomonas* sp. HR199 *vdh*ΩKm with *E. coli* S17-1 (pSUP*ech*ΩGm). The replacement of the intact *ech* gene with the ΩGm-inactivated gene (*ech*ΩGm) was verified by means of DNA sequencing.

Example 7

Generating of mutants of the strain *Pseudomonas* sp. HR199 in which genes of eugenol catabolism have been specifically inactivated by deleting a constituent region.

The strains *Pseudomonas* sp. HR199 *fcs*ΩKm, *Pseudomonas* sp. HR199 *ech*ΩKm, *Pseudomonas* sp. HR199 *vdh*ΩKm and *Pseudomonas* sp. HR199 *aat*ΩKm were employed as recipients in conjugation experiments in which strains of *E.* coli S17-1 harbouring the hybrid plasmids of pHE55 which are listed below were used as donors. The "heterogenotic" transconjugants were selected on gluconate-containing mineral medium which also contained the antibiotic corresponding to the Ω element in addition to tetracycline (pHE55-encoded resistance). After streaking out on sucrose-containing mineral medium, transconjugants were obtained which had eliminated the vector DNA by means of a second recombination event (second crossover). By streaking out on mineral medium which was without antibiotic or which contained the antibiotic corresponding to the Ω element, it was possible to identify the mutants in which the Ω element-inactivated gene had been replaced with the deletion-inactivated gene (no antibiotic resistance).

The mutant *Pseudomonas* sp. HR199 *fcs*Δ was obtained after conjugating *Pseudomonas* sp. HR199 *fcs*ΩKm with *E. coli* S17-1 (pHE*fcs*Δ). The replacement of the ΩKm inactivated gene (*fcs*ΩKm) with the deletion-inactivated gene (*fcs*Δ) was verified by means of DNA sequencing.

The mutant *Pseudomonas* sp. HR199 *ech*Δ was obtained after conjugating *Pseudomonas* sp. HR199 *ech*ΩKm with *E. coli* S17-1 (pHE*ech*Δ). The replacement of the ΩKm-inactivated gene (*ech*ΩKm) with the deletion-inactivated gene (*ech*Δ) was verified by means of DNA sequencing.

The mutant *Pseudomonas* sp. HR199 *vdh*Δ was obtained after conjugating *Pseudomonas* sp. HR199 *vdh*ΩKm with *E. coli* S17-1 (pHE*vdh*Δ). The replacement

of the ΩKm-inactivated gene (*vdh*ΩKm) with the deletion-inactivated gene (*vdh*Δ) was verified by means of DNA sequencing.

The mutant *Pseudomonas* sp. HR199 *aat*Δ was obtained after conjugating *Pseudomonas* sp. HR199 *aat*ΩKm with *E. coli* S17-1 (pHE*aat*Δ). The replacement of the ΩKm-inactivated gene (*aat*ΩKm) with the deletion-inactivated gene (*aat*Δ) was verified by means of DNA sequencing.

Example 8

Biotransforming eugenol into vanillin using the mutant *Pseudomonas* sp. HR199 *vdh*ΩKm.

The strain *Pseudomonas* sp. HR199 *vdh*ΩKm was propagated in 50 ml of HR-MM containing 6 mM eugenol up to an optical density of approx. OD600nm = 0.6. After 17 h, it was possible to detect 2.9 mM vanillin, 1.4 mM ferulic acid and 0.4 mM vanillic acid in the culture supernatant.

Example 9

Biotransforming eugenol into ferulic acid using the mutant *Pseudomonas* sp. HR199 *vdh*ΩGm*aat*ΩKm.

The strain *Pseudomonas* sp. HR199 *vdh*ΩGm*aat*ΩKm was propagated in 50 ml of HR-MM containing 6 mM eugenol up to an optical density of approx.OD600nm = 0.6. After 18 h, it was possible to detect 1.9 mM vanillin, 2.4 mM ferulic acid and 0.6 mM vanillic acid in the culture supernatant.

Example 10

Biotransforming eugenol into coniferyl alcohol using the mutant *Pseudomonas* sp. HR199 *vdh*ΩGm*aat*ΩKm.

The strain *Pseudomonas* sp. HR199 *vdh*ΩGm*aat*ΩKm was propagated in 50 ml of HR-MM containing 6 mM eugenol up to an optical density of approx. OD600nm = 0.4. After 15 h, it was possible to detect 1.7 mM coniferyl alcohol, 1.4 mM vanillin, 1.4 mM ferulic acid and 0.2 mM vanillic acid in the culture supernatant.

Example 11

Fermentatively producing natural vanillin from eugenol in a 10 l fermenter using mutant Pseudomonas sp. HR 199 *vdh*ΩKm.

The production fermenter was inoculated with 100 ml of a 24-hour-old preliminary culture which had been propagated at 32°C on a shaking incubator (120 rpm) in a medium which was adjusted to pH 7.0 and which consisted of 12.5 g of glycerol/l, 10 g of yeast extract/l and 0.37 g of acetic acid/l. The fermenter contained 9.9 l of medium of the following composition: 1.5 g of yeast extract/l, 1.6 g of KH_2PO_4/l, 0.2 g of NaCl/l, 0.2 g of $MgSO_4$/l. The pH was adjusted to pH 7.0 with sodium hydroxide solution. After sterilization, 4 g of eugenol were added to the medium. The temperature was 32°C, the aeration was 3 Nl/min and the stirrer speed was 600 rpm. The pH was maintained at pH 6.5 with sodium hydroxide solution.

At 4 hours after the inoculation, continuous addition of eugenol was begun such that 255 g of eugenol had been added to the culture when fermentation ended after 65 hours. 40 g of yeast extract were also fed in during the fermentation. At the end of the fermentation, the concentration of eugenol was 0.2 g/l. The content of vanillin was 2.6 g/l. 3.4 g of ferulic acid/l were also present.

The vanillin which is obtained in this way can be isolated by known physical methods such as chromatography, distillation and/or extraction and used for preparing natural flavourings.

Explanatory notes regarding the figures:

FIG. 1a to 1r:

Gene struktures for isolating organisms and mutants

*calA**: Part of the inactivated gene for coniferyl alcohol dehydrogenase
*calB**: Part of the inactivated gene for coniferyl aldehyde dehydrogenase
*fcs**: Part of the inactivated gene for feruloyl-CoA synthetase
*ech**: Part of the inactivated gene for enoyl-CoA hydratase-aldolase
*vdh**: Part of the inactivated gene for vanillin dehydrogenase
*aat**: Part of the inactivated gene for beta-ketothiolase

While the restriction enzyme cleavage sites labelled "*" were used for the construction, they are no longer functional in the resulting construct.

FIG. 2a: Nucleotide sequence of the *calA*ΩKm gene structure

FIG. 2b: Nucleotide sequence of the *calA*ΩGm gene structure:

FIG. 2c: Nucleotide sequence of the *calA*Δ gene structure

FIG. 2d: Nucleotide sequence of the *calB*ΩKm gene structure

FIG. 2e: Nucleotide sequence of the *calB*ΩGm gene structure

FIG. 2f: Nucleotide sequence of the *calB*Δ gene structure

FIG. 2g: Nucleotide sequence of the *fcs*ΩKm gene structure

FIG. 2h: Nucleotide sequence of the *fcs*ΩGm gene structure

FIG. 2i: Nucleotide sequence of the *fcs*Δ gene structure

FIG. 2j: Nucleotide sequence of the *ech*ΩKm gene structure

FIG. 2k: Nucleotide sequence of the *ech*ΩGm gene structure

FIG. 2l: Nucleotide sequence of the *ech*Δ gene structure

FIG. 2m: Nucleotide sequence of the *vdh*ΩKm gene structure

FIG. 2n: Nucleotide sequence of the *vdh*ΩGm gene structure

FIG. 2o: Nucleotide sequence of the vdhΔ gene structure

FIG. 2p: Nucleotide sequence of the *aat*ΩKm gene structure

FIG. 2q: Nucleotide sequence of the *aat*ΩGm gene structure

FIG. 2r: Nucleotide sequence of the *aat*Δ gene structure

Patent claims

1. Transformed and/or mutagenized unicellular or multicellular organism which is characterized in that enzymes of eugenol and/or ferulic acid catabolism are inactivated such that the intermediates coniferyl alcohol, coniferyl aldehyde, ferulic acid, vanillin and/or vanillic acid accumulate.

2. Organism according to Claim 1, characterized in that eugenol and/or ferulic acid catabolism is altered by inserting Ω elements, or introducing deletions, into corresponding genes.

3. Organism according to either Claim 1 or 2, characterized in that one or more genes encoding the enzymes coniferyl alcohol dehydrogenases, coniferyl aldehyde dehydrogenases, feruloyl-CoA synthetases, enoyl-CoA hydratase-aldolases, beta-ketothiolases, vanillin dehydrogenases or vanillic acid demethylases is/are altered and/or inactivated.

4. Organism according to one of Claims 1 to 3, characterized in that it is unicellular, preferably a microorganism or a plant or animal cell.

5. Organism according to one of Claims 1 to 4, characterized in that it is a bacterium, preferably a *Pseudomonas* species.

6. Gene structures in which the nucleotide sequences encoding the enzymes coniferyl alcohol dehydrogenases, coniferyl aldehyde dehydrogenases, feruloyl-CoA synthetases, enoyl-CoA hydratase-aldolases, beta-ketothiolases, vanillin-dehydrogenases or vanillic acid demethylases, or two or more of these enzymes, are altered and/or inactivated.

7. Gene structures having the sequences given in Figures 1a to 1r.

8. Gene structures having the sequences given in Figures 2a to 2r.

9. Vectors which contain at least one gene structure according to one of Claims 6 to 8.

10. Transformed organism according to one of Claims 1 to 5, characterized in that it harbours at least one vector according to Claim 9.

11. Organism according to one of Claims 1 to 5, characterized in that it contains at least one gene structure according to one of Claims 6 to 8 which is integrated into the genome instead of the respective intact gene.

12. Process for the biotechnological preparation of organic compounds, in particular alcohols, aldehydes and organic acids, characterized in that an organism according to one of Claims 1 to 5 or 10 to 11 is employed.

13. Process for preparing the organisms according to one of Claims 1 to 5, characterized in that the alteration eugenol and/or ferulic acid catabolism is achieved by means of microbiological culturing methods which are known per se.

14. Process for preparing an organism according to one of Claims 1 to 5 or 10 to 11, characterized in that the alteration in eugenol and/or ferulic acid catabolism, and/or the inactivation of the corresponding genes, is achieved by means of recombinant DNA methods.

15. Use of the organisms according to one of Claims 1 to 5 or 10 to 11 for preparing coniferyl alcohol, coniferyl aldehyde, ferulic acid, vanillin and/or vanillic acid.

16. Use of gene structures according to one of Claims 6 to 8 or of a vector according to Claim 9 for preparing transformed and/or mutagenized organisms.

Sequences

```
CTGCAGCCAG GGCTGAAAAG GAGGGATTCA GTGAGGTCAT GAAGGGAGGG GACGGCGCCT      60

GGCTCCAATT GCTCGATGGC GCCGCGATTG AGTGTCTTGG GCGCGGTCTT GGAGAGTTCG     120

GCTAGGGAGA TAAATTTGCT GGCCATGGTG GCGGCCCCTG ATGGGTTGGA TGATTTTCTG     180

CATTCTGCAT CATGAAATTC ATGAAATCAT CACTTTTCGG GGGGTGGGTG CACGGGATTG     240

AAGGTTGCTA GGAGAGTGCA TTGCTCGTAA GCCCAGGAAG CACGCGGGTT TCAGGATGGT     300

GCATGGAAAT GGCATGAGCT TTGCTGGATA TGATTAGAGA CATTAACTAT TTTGGCGGAA     360

TGGAAGCACG ATTCCTCGCC CGGTAGAGCG GTAACCGCGA CATTCAGGAC CGTAAAAAGG     420

AAAGAGCATG CAA CTG ACC AAC AAG AAA ATC GTC GTC ACC GGA GTG TCC TCC     472
       Met Gln Leu Thr Asn Lys Lys Ile Val Val Thr Gly Val Ser Ser
         1               5                  10                  15

GGT ATC GGT GCC GAA ACT GCC CGC GTT CTG CGC TCT CAC GGC GCC ACA       520
Gly Ile Gly Ala Glu Thr Ala Arg Val Leu Arg Ser His Gly Ala Thr
                 20                  25                  30

GTG ATT GGC GTA GAT CGC AAC ATG CCG AGC CTG ACT CTG GAT GCT TTC       568
Val Ile Gly Val Asp Arg Asn Met Pro Ser Leu Thr Leu Asp Ala Phe
             35                  40                  45

GTT CAG GCT GAC CTG AGC CAT CCT GAA GGC ATC GAT AAG GCC ATC GGG       616
Val Gln Ala Asp Leu Ser His Pro Glu Gly Ile Asp Lys Ala Ile
         50                  55                  60      62

ACAGCAAGCG AACCGGAATT GCCAGCTGGG GCGCCCTCTG GTAAGGTTGG GAAGCCCTGC     676

AAAGTAAACT GGATGGCTTT CTTGCCGCCA AGGATCTGAT GGCGCAGGGG ATCAAGATCT     736

GATCAAGAGA CAGGATGAGG ATCGTTTCGC ATG ATT GAA CAA GAT GGA TTG CAC       790
                                 Met Ile Glu Gln Asp Gly Leu His
                                   1               5

GCA GGT TCT CCG GCC GCT TGG GTG GAG AGG CTA TTC GGC TAT GAC TGG       838
Ala Gly Ser Pro Ala Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp
     10                  15                  20

GCA CAA CAG ACA ATC GGC TGC TCT GAT GCC GCC GTG TTC CGG CTG TCA       886
Ala Gln Gln Thr Ile Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser
 25                  30                  35                  40

GCG CAG GGG CGC CCG GTT CTT TTT GTC AAG ACC GAC CTG TCC GGT GCC       934
Ala Gln Gly Arg Pro Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala
                 45                  50                  55

CTG AAT GAA CTG CAG GAC GAG GCA GCG CGG CTA TCG TGG CTG GCC ACG       982
Leu Asn Glu Leu Gln Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr
             60                  65                  70
```

```
ACG GGC GTT CCT TGC GCA GCT GTG CTC GAC GTT GTC ACT GAA GCG GGA        1030
Thr Gly Val Pro Cys Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly
         75                  80                  85

AGG GAC TGG CTG CTA TTG GGC GAA GTG CCG GGG CAG GAT CTC CTG TCA        1078
Arg Asp Trp Leu Leu Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser
     90                  95                 100

TCT CAC CTT GCT CCT GCC GAG AAA GTA TCC ATC ATG GCT GAT GCA ATG        1126
Ser His Leu Ala Pro Ala Glu Lys Val Ser Ile Met Ala Asp Ala Met
105                 110                 115                 120

CGG CGG CTG CAT ACG CTT GAT CCG GCT ACC TGC CCA TTC GAC CAC CAA        1174
Arg Arg Leu His Thr Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln
                125                 130                 135

GCG AAA CAT CGC ATC GAG CGA GCA CGT ACT CGG ATG GAA GCC GGT CTT        1222
Ala Lys His Arg Ile Glu Arg Ala Arg Thr Arg Met Glu Ala Gly Leu
            140                 145                 150

GTC GAT CAG GAT GAT CTG GAC GAA GAG CAT CAG GGG CTC GCG CCA GCC        1270
Val Asp Gln Asp Asp Leu Asp Glu Glu His Gln Gly Leu Ala Pro Ala
        155                 160                 165

GAA CTG TTC GCC AGG CTC AAG GCG CGC ATG CCC GAC GGC GAG GAT CTC        1318
Glu Leu Phe Ala Arg Leu Lys Ala Arg Met Pro Asp Gly Glu Asp Leu
    170                 175                 180

GTC GTG ACC CAT GGC GAT GCC TGC TTG CCG AAT ATC ATG GTG GAA AAT        1366
Val Val Thr His Gly Asp Ala Cys Leu Pro Asn Ile Met Val Glu Asn
185                 190                 195                 200

GGC CGC TTT TCT GGA TTC ATC GAC TGT GGC CGG CTG GGT GTG GCG GAC        1414
Gly Arg Phe Ser Gly Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp
                205                 210                 215

CGC TAT CAG GAC ATA GCG TTG GCT ACC CGT GAT ATT GCT GAA GAG CTT        1462
Arg Tyr Gln Asp Ile Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu
            220                 225                 230

GGC GGC GAA TGG GCT GAC CGC TTC CTC GTG CTT TAC GGT ATC GCC GCT        1510
Gly Gly Glu Trp Ala Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala
        235                 240                 245

CCC GAT TCG CAG CGC ATC GCC TTC TAT CGC CTT CTT GAC GAG TTC TTC        1558
Pro Asp Ser Gln Arg Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe
    250                 255                 260                 264

TGAGCGGGAC TCTGGGGTTC GAAATGACCG ACCAAGCGAC GCCCTG GCC GCG GTG         1613
                                                   Ala Ala Val
                                                   225

ATT GCA TTC ATG TGT GCT GAG GAG TCA CGT TGG ATC AAC GGC ATA AAT        1661
Ile Ala Phe Met Cys Ala Glu Glu Ser Arg Trp Ile Asn Gly Ile Asn
        230                 235                 240
```

```
ATT CCA GTG GAC GGA GGT TTG GCA TCG ACC TAC GTG TAA GTTCGTGGAC        1710
Ile Pro Val Asp Gly Gly Leu Ala Ser Thr Tyr Val
    245                 250                 255

GCCCTTTGCA CGCGCACTAT ATCTCTATGC AGCAGCTGAA AGCAGCTTTG GTTTTGATCG    1770

GAGGTAGCGG GCGGAAAGGT GCAGAATGTC TAAATAATAA AGGATTCTTG TGAAGCTTTA    1830

GTTGTCCGTA AACGAAAATA AAAATAAAGA GGAATGATAT GAAAGCAAGT AGATCAGTCT    1890

GCACTTTCAA AATAGCTACC CTGGCAGGCG CCATTTATGC AGCGCTGCCA ATGTCAGCTG    1950

CAAACTCGAT GCAGCTGGAT GTAGGTAGCT CGGATTGGAC GGTGCGTTGG GGACAACACC    2010

CTCAAGTATA GCCTTGCCTC TCGCCTGAAT GAGCAAGACT CAAGTCTGAC AAATGCGCCG    2070

ACTGTCAATG GTTATATCCG GATATTCAAA GTCAGGGTGA TCGTAACTTT GACCGGGGGC    2130

TTGGTATCCA ATCGTCTCGA TATTCTGGCT GCAG                                2164
```

FIG. 2a:

```
CTGCAGCCAG GGCTGAAAAG GAGGGATTCA GTGAGGTCAT GAAGGGAGGG GACGGCGCCT      60

GGCTCCAATT GCTCGATGGC GCCGCGATTG AGTGTCTTGG GCGCGGTCTT GGAGAGTTCG     120

GCTAGGGAGA TAAATTTGCT GGCCATGGTG GCGGCCCCTG ATGGGTTGGA TGATTTTCTG     180

CATTCTGCAT CATGAAATTC ATGAAATCAT CACTTTTCGG GGGGTGGGTG CACGGGATTG     240

AAGGTTGCTA GGAGAGTGCA TTGCTCGTAA GCCCAGGAAG CACGCGGGTT TCAGGATGGT     300

GCATGGAAAT GGCATGAGCT TTGCTGGATA TGATTAGAGA CATTAACTAT TTTGGCGGAA     360

TGGAAGCACG ATTCCTCGCC CGGTAGAGCG GTAACCGCGA CATTCAGGAC CGTAAAAAGG     420

AAAGAGCATG CAA CTG ACC AAC AAG AAA ATC GTC GTC ACC GGA GTG TCC TCC     472
       Met Gln Leu Thr Asn Lys Lys Ile Val Val Thr Gly Val Ser Ser
         1               5                  10                  15

GGT ATC GGT GCC GAA ACT GCC CGC GTT CTG CGC TCT CAC GGC GCC ACA       520
Gly Ile Gly Ala Glu Thr Ala Arg Val Leu Arg Ser His Gly Ala Thr
                20                  25                  30

GTG ATT GGC GTA GAT CGC AAC ATG CCG AGC CTG ACT CTG GAT GCT TTC       568
Val Ile Gly Val Asp Arg Asn Met Pro Ser Leu Thr Leu Asp Ala Phe
            35                  40                  45

GTT CAG GCT GAC CTG AGC CAT CCT GAGGGGAGAG GCGGTTTGCG TATTGGGCGC      622
Val Gln Ala Asp Leu Ser His Pro
        50                  55

ATGCATAAAA ACTGTTGTAA TTCATTAAGC ATTCTGCCGA CATGGAAGCC ATCACAAACG     682

GCATGATGAA CCTGAATCGC CAGCGGCATC AGCACCTTGT CGCCTTGCGT ATAATATTTG     742

CCCATGGACG CACACCGTGG AAACGGATGA AGGCACGAAC CCAGTTGACA TAAGCCTGTT     802

CGGTTCGTAA ACTGTAATGC AAGTAGCGTA TGCGCTCACG CAACTGGTCC AGAACCTTGA     862

CCGAACGCAG CGGTGGTAAC GGCGCAGTGG CGGTTTTCAT GGCTTGTTAT GACTGTTTTT     922

TTGTACAGTC TATGCCTCGG GCATCCAAGC AGCAAGCGCG TTACGCCGTG GGTCGATGTT     982

TGATGTTATG GAGCAGCAAC G ATG TTA CGC AGC AGC AAC GAT GTT ACG CAG       1033
                       Met Leu Arg Ser Ser Asn Asp Val Thr Gln
                         1               5                  10

CAG GGC AGT CGC CCT AAA ACA AAG TTA GGT GGC TCA AGT ATG GGC ATC       1081
Gln Gly Ser Arg Pro Lys Thr Lys Leu Gly Gly Ser Ser Met Gly Ile
                15                  20                  25

ATT CGC ACA TGT AGG CTC GGC CCT GAC CAA GTC AAA TCC ATG CGG GCT       1129
Ile Arg Thr Cys Arg Leu Gly Pro Asp Gln Val Lys Ser Met Arg Ala
            30                  35                  40

GCT CTT GAT CTT TTC GGT CGT GAG TTC GGA GAC GTA GCC ACC TAC TCC       1177
Ala Leu Asp Leu Phe Gly Arg Glu Phe Gly Asp Val Ala Thr Tyr Ser
        45                  50                  55
```

```
CAA CAT CAG CCG GAC TCC GAT TAC CTC GGG AAC TTG CTC CGT AGT AAG      1225
Gln His Gln Pro Asp Ser Asp Tyr Leu Gly Asn Leu Leu Arg Ser Lys
     60                  65                  70

ACA TTC ATC GCG CTT GCT GCC TTC GAC CAA GAA GCG GTT GTT GGC GCT      1273
Thr Phe Ile Ala Leu Ala Ala Phe Asp Gln Glu Ala Val Val Gly Ala
 75                  80                  85                  90

CTC GCG GCT TAC GTT CTG CCC AGG TTT GAG CAG CCG CGT AGT GAG ATC      1321
Leu Ala Ala Tyr Val Leu Pro Arg Phe Glu Gln Pro Arg Ser Glu Ile
                95                  100                 105

TAT ATC TAT GAT CTC GCA GTC TCC GGC GAG CAC CGG AGG CAG GGC ATT      1369
Tyr Ile Tyr Asp Leu Ala Val Ser Gly Glu His Arg Arg Gln Gly Ile
            110                 115                 120

GCC ACC GCG CTC ATC AAT CTC CTC AAG CAT GAG GCC AAC GCG CTT GGT      1417
Ala Thr Ala Leu Ile Asn Leu Leu Lys His Glu Ala Asn Ala Leu Gly
        125                 130                 135

GCT TAT GTG ATC TAC GTG CAA GCA GAT TAC GGT GAC GAT CCC GCA GTG      1465
Ala Tyr Val Ile Tyr Val Gln Ala Asp Tyr Gly Asp Asp Pro Ala Val
    140                 145                 150

GCT CTC TAT ACA AAG TTG GGC ATA CGG GAA GAA GTG ATG CAC TTT GAT      1513
Ala Leu Tyr Thr Lys Leu Gly Ile Arg Glu Glu Val Met His Phe Asp
155                 160                 165                 170

ATC GAC CCA AGT ACC GCC ACC TAA CAATTCGTTC AAGCCGAGAT CGGCTTCCCT      1567
Ile Asp Pro Ser Thr Ala Thr
                175     177

G ATT GCA TTC ATG TGT GCT GAG GAG TCA CGT TGG ATC AAC GGC ATA AAT     1616
  Ile Ala Phe Met Cys Ala Glu Glu Ser Arg Trp Ile Asn Gly Ile Asn
  228     230                 235                 240

ATT CCA GTG GAC GGA GGT TTG GCA TCG ACC TAC GTG TAA GTTCGTGGAC       1665
Ile Pro Val Asp Gly Gly Leu Ala Ser Thr Tyr Val
    245                 250                 255

GCCCTTTGCA CGCGCACTAT ATCTCTATGC AGCAGCTGAA AGCAGCTTTG GTTTTGATCG    1725

GAGGTAGCGG GCGGAAAGGT GCAGAATGTC TAAATAATAA AGGATTCTTG TGAAGCTTTA    1785

GTTGTCCGTA AACGAAAATA AAAATAAAGA GGAATGATAT GAAAGCAAGT AGATCAGTCT    1845

GCACTTTCAA AATAGCTACC CTGGCAGGCG CCATTTATGC AGCGCTGCCA ATGTCAGCTG    1905

CAAACTCGAT GCAGCTGGAT GTAGGTAGCT CGGATTGGAC GGTGCGTTGG GGACAACACC    1965

CTCAAGTATA GCCTTGCCTC TCGCCTGAAT GAGCAAGACT CAAGTCTGAC AAATGCGCCG    2025

ACTGTCAATG GTTATATCCG GATATTCAAA GTCAGGGTGA TCGTAACTTT GACCGGGGGC    2085

TTGGTATCCA ATCGTCTCGA TATTCTGGCT GCAG                                2119
```

FIG. 2b:

```
CTGCAGCCAG GGCTGAAAAG GAGGGATTCA GTGAGGTCAT GAAGGGAGGG GACGGCGCCT      60

GGCTCCAATT GCTCGATGGC GCCGCGATTG AGTGTCTTGG GCGCGGTCTT GGAGAGTTCG     120

GCTAGGGAGA TAAATTTGCT GGCCATGGTG GCGGCCCCTG ATGGGTTGGA TGATTTTCTG     180

CATTCTGCAT CATGAAATTC ATGAAATCAT CACTTTTCGG GGGGTGGGTG CACGGGATTG     240

AAGGTTGCTA GGAGAGTGCA TTGCTCGTAA GCCCAGGAAG CACGCGGGTT TCAGGATGGT     300

GCATGGAAAT GGCATGAGCT TTGCTGGATA TGATTAGAGA CATTAACTAT TTTGGCGGAA     360

TGGAAGCACG ATTCCTCGCC CGGTAGAGCG GTAACCGCGA CATTCAGGAC CGTAAAAAGG     420

AAAGAGCATG CAA CTG ACC AAC AAG AAA ATC GTC GTC ACC GGA GTG TCC TCC     472
       Met Gln Leu Thr Asn Lys Lys Ile Val Val Thr Gly Val Ser Ser
         1               5                  10                  15

GGT ATC GGT GCC GAA ACT GCC CGC GTT CTG CGC TCT CAC GGC GCC ACA       520
Gly Ile Gly Ala Glu Thr Ala Arg Val Leu Arg Ser His Gly Ala Thr
                20                  25                  30

GTG ATT GGC GTA GAT CGC AAC ATG CCG AGC CTG ACT CTG GAT GCT TTC       568
Val Ile Gly Val Asp Arg Asn Met Pro Ser Leu Thr Leu Asp Ala Phe
            35                  40                  45

GTT CAG GCT GAC CTG AGC CAT CCT GAA GGC ATC GATC AAC GGC ATA AAT      617
Val Gln Ala Asp Leu Ser His Pro Glu Gly Ile      Asn Gly Ile Asn
        50                  55          58      240

ATT CCA GTG GAC GGA GGT TTG GCA TCG ACC TAC GTG TAA GTTCGTGGAC        666
Ile Pro Val Asp Gly Gly Leu Ala Ser Thr Tyr Val
    245                 250                 255

GCCCTTTGCA CGCGCACTAT ATCTCTATGC AGCAGCTGAA AGCAGCTTTG GTTTTGATCG     726

GAGGTAGCGG GCGGAAAGGT GCAGAATGTC TAAATAATAA AGGATTCTTG TGAAGCTTTA     786

GTTGTCCGTA AACGAAAATA AAAATAAAGA GGAATGATAT GAAAGCAAGT AGATCAGTCT     846

GCACTTTCAA AATAGCTACC CTGGCAGGCG CCATTTATGC AGCGCTGCCA ATGTCAGCTG     906

CAAACTCGAT GCAGCTGGAT GTAGGTAGCT CGGATTGGAC GGTGCGTTGG GGACAACACC     966

CTCAAGTATA GCCTTGCCTC TCGCCTGAAT GAGCAAGACT CAAGTCTGAC AAATGCGCCG    1026

ACTGTCAATG GTTATATCCG GATATTCAAA GTCAGGGTGA TCGTAACTTT GACCGGGGGC    1086

TTGGTATCCA ATCGTCTCGA TATTCTGGCT GCAG                                1120
```

FIG. 2c:

```
GAATTCCGCG TATCGCCCGG TTCTATCAGC GGGCCGCTTT CGAAAGTCAT GGTGTTAGCC        60

GGTAGGGTCT TTTTCTTGGC CATGCTTGTT GCCTGAACCT TCGTTGACAT AGGGCAGAGG       120

TGCGTTTGCC GCTTCGCTTC GCGATGAACC GCATCGAGAT GCTGAGGTCA GGATTTTTCC       180

TTAACTCGCG TAAGCATTCT GTCATTTTTT TGGTGGCTTT GAACAGCCTG ATGAAAGGTG       240

GTCTCGCCCT TTGAGGCCGA TTCTTGGGCG CTTGGCGGCG TCGAAGCGAT GCTCCACTAC       300

CGATTAAGAT AATTAAAATA AGGAAACCGC ATGGTTTCTT ATGTGAATTT GTCTGGCATA       360

CTCCAGCTCA AGGGCAATTT TTGGGCTATT GGCTGAGCAG TTGCCTCTAT ATGGTTATTC       420

AGAATAACAA TTGACTCCTC AGGAGGTCAG CG ATG AGC ATT CTT GGT TTG AAT         473
                                    Met Ser Ile Leu Gly Leu Asn
                                     1               5

GGT GCC CCG GTC GGA GCT GAG CAG CTG GGC TCG GCT CTT GAT CGC ATG         521
Gly Ala Pro Val Gly Ala Glu Gln Leu Gly Ser Ala Leu Asp Arg Met
        10                  15                  20

AAG AAG GCG CAC CTG GAG CAG GGG CCT GCA AAC TTG GAG CTG CGT CTG         569
Lys Lys Ala His Leu Glu Gln Gly Pro Ala Asn Leu Glu Leu Arg Leu
    25                  30                  35

AGT AGG CTG GAT CGT GCG ATT GCA ATG CTT CTG GAA AAT CGT GAA GCA         617
Ser Arg Leu Asp Arg Ala Ile Ala Met Leu Leu Glu Asn Arg Glu Ala
40                  45                  50                  55

ATT GCC GAC GCG GTT TCT GCT GAC TTT GGC AAT CGC AGC CGT GAG CAA         665
Ile Ala Asp Ala Val Ser Ala Asp Phe Gly Asn Arg Ser Arg Glu Gln
                60                  65                  70

ACA CTG CTT TGC GAC ATT GCT GGC TCG GTG GCA AGC CTG AAG GAT AGC         713
Thr Leu Leu Cys Asp Ile Ala Gly Ser Val Ala Ser Leu Lys Asp Ser
            75                  80                  85

CGC GAG CAC GTG GCC AAA TGG ATG GAG CCC GAA CAT CAC AAG GCG ATG         761
Arg Glu His Val Ala Lys Trp Met Glu Pro Glu His His Lys Ala Met
        90                  95                  100

TTT CCA GGG GCG GAG GCA CGC GTT GAG TTT CAG CCG CTG GGT GTC GTT         809
Phe Pro Gly Ala Glu Ala Arg Val Glu Phe Gln Pro Leu Gly Val Val
    105                 110                 115

GGG GTC ATT AGT CCC TGG AAC TTC CCT ATC GTA CTG GCC TTT GGG CCG         857
Gly Val Ile Ser Pro Trp Asn Phe Pro Ile Val Leu Ala Phe Gly Pro
120                 125                 130                 135

CTG GCC GGC ATA TTC GCA GCA GGT AAT CGC GCC ATG CTC AAG CCG TCC         905
Leu Ala Gly Ile Phe Ala Ala Gly Asn Arg Ala Met Leu Lys Pro Ser
                140                 145                 150

GAG CTT ACC CCG CGG ACT TCT GCC CTG CTT GCG GAG CTA ATT GCT CGT         953
Glu Leu Thr Pro Arg Thr Ser Ala Leu Leu Ala Glu Leu Ile Ala Arg
            155                 160                 165
```

```
TAC TTC GAT GAA ACT GAG CTG ACT ACA GTG CTG GGC GAC GCT GAA GTC         1001
Tyr Phe Asp Glu Thr Glu Leu Thr Thr Val Leu Gly Asp Ala Glu Val
        170                 175                 180

GGT GCG CTG TTC AGT GCT CAG CCT TTC GAT CAT CTG ATC TTC ACC GGC         1049
Gly Ala Leu Phe Ser Ala Gln Pro Phe Asp His Leu Ile Phe Thr Gly
    185                 190                 195

GGC ACT GCC GTG GCC AAG CAC ATC ATG CGT GCC GCG GCG GAT AAC CTA         1097
Gly Thr Ala Val Ala Lys His Ile Met Arg Ala Ala Ala Asp Asn Leu
200                 205                 210                 215

GTG CCC GTT ACC CTG GAA TTG GGT GGC AAA TCG CCG GTG ATC GTT TCC         1145
Val Pro Val Thr Leu Glu Leu Gly Gly Lys Ser Pro Val Ile Val Ser
                220                 225                 230

CGC AGT GCA GAT ATG GCG GAC GTT GCA CAA CGG GTG TTG ACG GTG AAA         1193
Arg Ser Ala Asp Met Ala Asp Val Ala Gln Arg Val Leu Thr Val Lys
            235                 240                 245

ACC TTC AAT GCC GGG CAA ATC TGT CTG GCA CCG GAC TAT GTG CTG CTG         1241
Thr Phe Asn Ala Gly Gln Ile Cys Leu Ala Pro Asp Tyr Val Leu Leu
        250                 255                 260

CCG GAA GGGACAGCAA GCGAACCGGA ATTGCCAGCT GGGGCGCCCT CTGGTAAGGT          1297
Pro Glu
    265

TGGGAAGCCC TGCAAAGTAA ACTGGATGGC TTTCTTGCCG CCAAGGATCT GATGGCGCAG       1357

GGGATCAAGA TCTGATCAAG AGACAGGATG AGGATCGTTT CGC ATG ATT GAA CAA         1412
                                                Met Ile Glu Gln
                                                  1

GAT GGA TTG CAC GCA GGT TCT CCG GCC GCT TGG GTG GAG AGG CTA TTC         1460
Asp Gly Leu His Ala Gly Ser Pro Ala Ala Trp Val Glu Arg Leu Phe
  5                  10                  15                  20

GGC TAT GAC TGG GCA CAA CAG ACA ATC GGC TGC TCT GAT GCC GCC GTG         1508
Gly Tyr Asp Trp Ala Gln Gln Thr Ile Gly Cys Ser Asp Ala Ala Val
                25                  30                  35

TTC CGG CTG TCA GCG CAG GGG CGC CCG GTT CTT TTT GTC AAG ACC GAC         1556
Phe Arg Leu Ser Ala Gln Gly Arg Pro Val Leu Phe Val Lys Thr Asp
            40                  45                  50

CTG TCC GGT GCC CTG AAT GAA CTG CAG GAC GAG GCA GCG CGG CTA TCG         1604
Leu Ser Gly Ala Leu Asn Glu Leu Gln Asp Glu Ala Ala Arg Leu Ser
        55                  60                  65

TGG CTG GCC ACG ACG GGC GTT CCT TGC GCA GCT GTG CTC GAC GTT GTC         1652
Trp Leu Ala Thr Thr Gly Val Pro Cys Ala Ala Val Leu Asp Val Val
    70                  75                  80
```

```
ACT GAA GCG GGA AGG GAC TGG CTG CTA TTG GGC GAA GTG CCG GGG CAG      1700
Thr Glu Ala Gly Arg Asp Trp Leu Leu Leu Gly Glu Val Pro Gly Gln
 85                  90                  95                  100

GAT CTC CTG TCA TCT CAC CTT GCT CCT GCC GAG AAA GTA TCC ATC ATG      1748
Asp Leu Leu Ser Ser His Leu Ala Pro Ala Glu Lys Val Ser Ile Met
                105                 110                 115

GCT GAT GCA ATG CGG CGG CTG CAT ACG CTT GAT CCG GCT ACC TGC CCA      1796
Ala Asp Ala Met Arg Arg Leu His Thr Leu Asp Pro Ala Thr Cys Pro
            120                 125                 130

TTC GAC CAC CAA GCG AAA CAT CGC ATC GAG CGA GCA CGT ACT CGG ATG      1844
Phe Asp His Gln Ala Lys His Arg Ile Glu Arg Ala Arg Thr Arg Met
        135                 140                 145

GAA GCC GGT CTT GTC GAT CAG GAT GAT CTG GAC GAA GAG CAT CAG GGG      1892
Glu Ala Gly Leu Val Asp Gln Asp Asp Leu Asp Glu Glu His Gln Gly
    150                 155                 160

CTC GCG CCA GCC GAA CTG TTC GCC AGG CTC AAG GCG CGC ATG CCC GAC      1940
Leu Ala Pro Ala Glu Leu Phe Ala Arg Leu Lys Ala Arg Met Pro Asp
165                 170                 175                 180

GGC GAG GAT CTC GTC GTG ACC CAT GGC GAT GCC TGC TTG CCG AAT ATC      1988
Gly Glu Asp Leu Val Val Thr His Gly Asp Ala Cys Leu Pro Asn Ile
                185                 190                 195

ATG GTG GAA AAT GGC CGC TTT TCT GGA TTC ATC GAC TGT GGC CGG CTG      2036
Met Val Glu Asn Gly Arg Phe Ser Gly Phe Ile Asp Cys Gly Arg Leu
            200                 205                 210

GGT GTG GCG GAC CGC TAT CAG GAC ATA GCG TTG GCT ACC CGT GAT ATT      2084
Gly Val Ala Asp Arg Tyr Gln Asp Ile Ala Leu Ala Thr Arg Asp Ile
        215                 220                 225

GCT GAA GAG CTT GGC GGC GAA TGG GCT GAC CGC TTC CTC GTG CTT TAC      2132
Ala Glu Glu Leu Gly Gly Glu Trp Ala Asp Arg Phe Leu Val Leu Tyr
    230                 235                 240

GGT ATC GCC GCT CCC GAT TCG CAG CGC ATC GCC TTC TAT CGC CTT CTT      2180
Gly Ile Ala Ala Pro Asp Ser Gln Arg Ile Ala Phe Tyr Arg Leu Leu
245                 250                 255                 260

GAC GAG TTC TTC TGA GCGGGACTCT GGGGTTCGAA ATGACCGACC AAGCGACGCC      2235
Asp Glu Phe Phe
            264

CGC CAT GCC AAG CCT GTT CTC GTG CAA AGT CCT GTG GGT GAG TCG AAC      2283
    His Ala Lys Pro Val Leu Val Gln Ser Pro Val Gly Glu Ser Asn
    444 445                 450                 455

TTG GCG ATG CGC GCA CCC TAC GGA GAA GCG ATC CAC GGA CTG CTC TCT      2331
Leu Ala Met Arg Ala Pro Tyr Gly Glu Ala Ile His Gly Leu Leu Ser
    460                 465                 470
```

```
GTC CTC CTT TCA ACG GAG TGT TAG AACCGTTGGT AGTGGTTTTG GACGGGCCCA     2385
Val Leu Leu Ser Thr Glu Cys
475                 480 481

GGAGCATGCG CTTCTGGGCC CGTTTCTTGA GTATTCATTG GATAGTCACG CGTGGTAGCT     2445

TCGAGCCTGC ACAGCTGATG AGCACCCTGG AAGGCGCGCT GTACGCGGAC GACTGGGTTC     2505

ATCTTCGCCA TTCATGACGG AACTCCGTTC CCCAGTACCG CGATGACTAT TTTGCCTCTT     2565

CCGATGTCCG ATTCCACGCC GCCTGACGCT AAGCGGGGGC GGGGGCGCCC GCATCCCAGC     2625

CCAGACAGCA ACAAATGAGT AGGCTCTTGG ATGCCGCGGC GGCTGAGATT GGTAACGGCA     2685

ATTTCGTCAA TGTGACGATG GATTCGATTG CCCGTGCTGC CGGCGTCTCA AAAAAAACGC     2745

TGTACGTCTT GGTGGCGAGC AAGGAAGAAC TCATTTCCCG GTTAGTGGCT CGAGACATGT     2805

CCAACCTTGA GGAATTC                                                    2822
```

FIG. 2d:

```
GAATTCCGCG TATCGCCCGG TTCTATCAGC GGGCCGCTTT CGAAAGTCAT GGTGTTAGCC      60

GGTAGGGTCT TTTTCTTGGC CATGCTTGTT GCCTGAACCT TCGTTGACAT AGGGCAGAGG     120

TGCGTTTGCC GCTTCGCTTC GCGATGAACC GCATCGAGAT GCTGAGGTCA GGATTTTTCC     180

TTAACTCGCG TAAGCATTCT GTCATTTTTT TGGTGGCTTT GAACAGCCTG ATGAAAGGTG     240

GTCTCGCCCT TTGAGGCCGA TTCTTGGGCG CTTGGCGGCG TCGAAGCGAT GCTCCACTAC     300

CGATTAAGAT AATTAAAATA AGGAAACCGC ATGGTTTCTT ATGTGAATTT GTCTGGCATA     360

CTCCAGCTCA AGGGCAATTT TTGGGCTATT GGCTGAGCAG TTGCCTCTAT ATGGTTATTC     420

AGAATAACAA TTGACTCCTC AGGAGGTCAG CG ATG AGC ATT CTT GGT TTG AAT        473
                                   Met Ser Ile Leu Gly Leu Asn
                                     1               5

GGT GCC CCG GTC GGA GCT GAG CAG CTG GGC TCG GCT CTT GAT CGC ATG        521
Gly Ala Pro Val Gly Ala Glu Gln Leu Gly Ser Ala Leu Asp Arg Met
         10                  15                  20

AAG AAG GCG CAC CTG GAG CAG GGG CCT GCA AAC TTG GAG CTG CGT CTG        569
Lys Lys Ala His Leu Glu Gln Gly Pro Ala Asn Leu Glu Leu Arg Leu
     25                  30                  35

AGT AGG CTG GAT CGT GCG ATT GCA ATG CTT CTG GAA AAT CGT GAA GCA        617
Ser Arg Leu Asp Arg Ala Ile Ala Met Leu Leu Glu Asn Arg Glu Ala
 40                  45                  50                  55

ATT GCC GAC GCG GTT TCT GCT GAC TTT GGC AAT CGC AGC CGT GAG CAA        665
Ile Ala Asp Ala Val Ser Ala Asp Phe Gly Asn Arg Ser Arg Glu Gln
                 60                  65                  70

ACA CTG CTT TGC GAC ATT GCT GGC TCG GTG GCA AGC CTG AAG GAT AGC        713
Thr Leu Leu Cys Asp Ile Ala Gly Ser Val Ala Ser Leu Lys Asp Ser
             75                  80                  85

CGC GAG CAC GTG GCC AAA TGG ATG GAG CCC GAA CAT CAC AAG GCG ATG        761
Arg Glu His Val Ala Lys Trp Met Glu Pro Glu His His Lys Ala Met
         90                  95                  100

TTT CCA GGG GCG GAG GCA CGC GTT GAG TTT CAG CCG CTG GGT GTC GTT        809
Phe Pro Gly Ala Glu Ala Arg Val Glu Phe Gln Pro Leu Gly Val Val
    105                 110                 115

GGG GTC ATT AGT CCC TGG AAC TTC CCT ATC GTA CTG GCC TTT GGG CCG        857
Gly Val Ile Ser Pro Trp Asn Phe Pro Ile Val Leu Ala Phe Gly Pro
120                 125                 130                 135

CTG GCC GGC ATA TTC GCA GCA GGT AAT CGC GCC ATG CTC AAG CCG TCC        905
Leu Ala Gly Ile Phe Ala Ala Gly Asn Arg Ala Met Leu Lys Pro Ser
                140                 145                 150

GAG CTT ACC CCG CGG ACT TCT GCC CTG CTT GCG GAG CTA ATT GCT CGT        953
Glu Leu Thr Pro Arg Thr Ser Ala Leu Leu Ala Glu Leu Ile Ala Arg
            155                 160                 165
```

```
TAC TTC GAT GAA ACT GAG CTG ACT ACA GTG CTG GGC GAC GCT GAA GTC      1001
Tyr Phe Asp Glu Thr Glu Leu Thr Thr Val Leu Gly Asp Ala Glu Val
       170                 175                 180

GGT GCG CTG TTC AGT GCT CAG CCT TTC GAT CAT CTG ATC TTC ACC GGC      1049
Gly Ala Leu Phe Ser Ala Gln Pro Phe Asp His Leu Ile Phe Thr Gly
    185                 190                 195

GGC ACT GCC GTG GCC AAG CAC ATC ATG CGT GCC GCG GCG GAT AAC CTA      1097
Gly Thr Ala Val Ala Lys His Ile Met Arg Ala Ala Ala Asp Asn Leu
200                 205                 210                 215

GTG CCC GTT ACC CTG GAA TTG GGT GGC AAA TCG CCG GTG ATC GTT TCC      1145
Val Pro Val Thr Leu Glu Leu Gly Gly Lys Ser Pro Val Ile Val Ser
                220                 225                 230

CGC AGT GCA GAT ATG GCG GAC GTT GCA CAA CGG GTG TTG ACG GTG AAA      1193
Arg Ser Ala Asp Met Ala Asp Val Ala Gln Arg Val Leu Thr Val Lys
            235                 240                 245

ACC TTC AAT GCC GGG CAA ATC TGT CTG GCA CCG GAC TAT GTG CTG GGG      1241
Thr Phe Asn Ala Gly Gln Ile Cys Leu Ala Pro Asp Tyr Val Leu
        250                 255                 260     262

GAGAGGCGGT TTGCGTATTG GGCGCATGCA TAAAAACTGT TGTAATTCAT TAAGCATTCT    1301

GCCGACATGG AAGCCATCAC AAACGGCATG ATGAACCTGA ATCGCCAGCG GCATCAGCAC    1361

CTTGTCGCCT TGCGTATAAT ATTTGCCCAT GGACGCACAC CGTGGAAACG GATGAAGGCA    1421

CGAACCCAGT TGACATAAGC CTGTTCGGTT CGTAAACTGT AATGCAAGTA GCGTATGCGC    1481

TCACGCAACT GGTCCAGAAC CTTGACCGAA CGCAGCGGTG GTAACGGCGC AGTGGCGGTT    1541

TTCATGGCTT GTTATGACTG TTTTTTTGTA CAGTCTATGC CTCGGGCATC CAAGCAGCAA    1601

GCGCGTTACG CCGTGGGTCG ATGTTTGATG TTATGGAGCA GCAACG ATG TTA CGC       1656
                                                    Met Leu Arg
                                                      1

AGC AGC AAC GAT GTT ACG CAG CAG GGC AGT CGC CCT AAA ACA AAG TTA      1704
Ser Ser Asn Asp Val Thr Gln Gln Gly Ser Arg Pro Lys Thr Lys Leu
      5                  10                  15

GGT GGC TCA AGT ATG GGC ATC ATT CGC ACA TGT AGG CTC GGC CCT GAC      1752
Gly Gly Ser Ser Met Gly Ile Ile Arg Thr Cys Arg Leu Gly Pro Asp
 20                  25                  30                  35

CAA GTC AAA TCC ATG CGG GCT GCT CTT GAT CTT TTC GGT CGT GAG TTC      1800
Gln Val Lys Ser Met Arg Ala Ala Leu Asp Leu Phe Gly Arg Glu Phe
                40                  45                  50

GGA GAC GTA GCC ACC TAC TCC CAA CAT CAG CCG GAC TCC GAT TAC CTC      1848
Gly Asp Val Ala Thr Tyr Ser Gln His Gln Pro Asp Ser Asp Tyr Leu
            55                  60                  65

GGG AAC TTG CTC CGT AGT AAG ACA TTC ATC GCG CTT GCT GCC TTC GAC      1896
Gly Asn Leu Leu Arg Ser Lys Thr Phe Ile Ala Leu Ala Ala Phe Asp
        70                  75                  80

CAA GAA GCG GTT GTT GGC GCT CTC GCG GCT TAC GTT CTG CCC AGG TTT      1944
```

```
Gln Glu Ala Val Val Gly Ala Leu Ala Ala Tyr Val Leu Pro Arg Phe
     85                  90                  95

GAG CAG CCG CGT AGT GAG ATC TAT ATC TAT GAT CTC GCA GTC TCC GGC      1992
Glu Gln Pro Arg Ser Glu Ile Tyr Ile Tyr Asp Leu Ala Val Ser Gly
100                 105                 110                 115

GAG CAC CGG AGG CAG GGC ATT GCC ACC GCG CTC ATC AAT CTC CTC AAG      2040
Glu His Arg Arg Gln Gly Ile Ala Thr Ala Leu Ile Asn Leu Leu Lys
                120                 125                 130

CAT GAG GCC AAC GCG CTT GGT GCT TAT GTG ATC TAC GTG CAA GCA GAT      2088
His Glu Ala Asn Ala Leu Gly Ala Tyr Val Ile Tyr Val Gln Ala Asp
            135                 140                 145

TAC GGT GAC GAT CCC GCA GTG GCT CTC TAT ACA AAG TTG GGC ATA CGG      2136
Tyr Gly Asp Asp Pro Ala Val Ala Leu Tyr Thr Lys Leu Gly Ile Arg
        150                 155                 160

GAA GAA GTG ATG CAC TTT GAT ATC GAC CCA AGT ACC GCC ACC TAA CAA      2184
Glu Glu Val Met His Phe Asp Ile Asp Pro Ser Thr Ala Thr
    165                 170                 175     177

TTCGTTCAAG CCGAGATCGG CTTCCCTG CAA AGT CCT GTG GGT GAG TCG AAC       2236
                              Gln Ser Pro Val Gly Glu Ser Asn
                              451             455

TTG GCG ATG CGC GCA CCC TAC GGA GAA GCG ATC CAC GGA CTG CTC TCT      2284
Leu Ala Met Arg Ala Pro Tyr Gly Glu Ala Ile His Gly Leu Leu Ser
    460                 465                 470

GTC CTC CTT TCA ACG GAG TGT TAG AACCGTTGGT AGTGGTTTTG GACGGGCCCA     2338
Val Leu Leu Ser Thr Glu Cys
475                 480 481

GGAGCATGCG CTTCTGGGCC CGTTTCTTGA GTATTCATTG GATAGTCACG CGTGGTAGCT    2398

TCGAGCCTGC ACAGCTGATG AGCACCCTGG AAGGCGCGCT GTACGCGGAC GACTGGGTTC    2458

ATCTTCGCCA TTCATGACGG AACTCCGTTC CCCAGTACCG CGATGACTAT TTTGCCTCTT    2518

CCGATGTCCG ATTCCACGCC GCCTGACGCT AAGCGGGGGC GGGGGCGCCC GCATCCCAGC    2578

CCAGACAGCA ACAAATGAGT AGGCTCTTGG ATGCCGCGGC GGCTGAGATT GGTAACGGCA    2638

ATTTCGTCAA TGTGACGATG GATTCGATTG CCCGTGCTGC CGGCGTCTCA AAAAAAACGC    2698

TGTACGTCTT GGTGGCGAGC AAGGAAGAAC TCATTTCCCG GTTAGTGGCT CGAGACATGT    2758

CCAACCTTGA GGAATTC                                                   2775
```

FIG. 2e:

```
GAATTCCGCG TATCGCCCGG TTCTATCAGC GGGCCGCTTT CGAAAGTCAT GGTGTTAGCC      60

GGTAGGGTCT TTTTCTTGGC CATGCTTGTT GCCTGAACCT TCGTTGACAT AGGGCAGAGG     120

TGCGTTTGCC GCTTCGCTTC GCGATGAACC GCATCGAGAT GCTGAGGTCA GGATTTTTCC     180

TTAACTCGCG TAAGCATTCT GTCATTTTTT TGGTGGCTTT GAACAGCCTG ATGAAAGGTG     240

GTCTCGCCCT TTGAGGCCGA TTCTTGGGCG CTTGGCGGCG TCGAAGCGAT GCTCCACTAC     300

CGATTAAGAT AATTAAAATA AGGAAACCGC ATGGTTTCTT ATGTGAATTT GTCTGGCATA     360

CTCCAGCTCA AGGGCAATTT TTGGGCTATT GGCTGAGCAG TTGCCTCTAT ATGGTTATTC     420

AGAATAACAA TTGACTCCTC AGGAGGTCAG CG ATG AGC ATT CTT GGT TTG AAT      473
                                    Met Ser Ile Leu Gly Leu Asn
                                     1               5

GGT GCC CCG GTC GGA GCT GAG CAG CTG GGC TCG GCT CTT GAT CGC ATG      521
Gly Ala Pro Val Gly Ala Glu Gln Leu Gly Ser Ala Leu Asp Arg Met
         10                  15                  20

AAG AAG GCG CAC CTG GAG CAG GGG CCT GCA AAC TTG GAG CTG CGT CTG      569
Lys Lys Ala His Leu Glu Gln Gly Pro Ala Asn Leu Glu Leu Arg Leu
     25                  30                  35

AGT AGG CTG GAT CGT GCG ATT GCA ATG CTT CTG GAA AAT CGT GAA GCA      617
Ser Arg Leu Asp Arg Ala Ile Ala Met Leu Leu Glu Asn Arg Glu Ala
 40                  45                  50                  55

ATT GCC GAC GCG GTT TCT GCT GAC TTT GGC AAT CGC AGC CGT GAG CAA      665
Ile Ala Asp Ala Val Ser Ala Asp Phe Gly Asn Arg Ser Arg Glu Gln
                 60                  65                  70

ACA CTG CTT TGC GAC ATT GCT GGC TCG GTG GCA AGC CTG AAG GAT AGC      713
Thr Leu Leu Cys Asp Ile Ala Gly Ser Val Ala Ser Leu Lys Asp Ser
             75                  80                  85

CGC GAG CAC GTG GCC AAA TGG ATG GAG CCC GAA CAT CAC AAG GCG ATG      761
Arg Glu His Val Ala Lys Trp Met Glu Pro Glu His His Lys Ala Met
         90                  95                 100

TTT CCA GGG GCG GAG GCA CGC GTT GAG TTT CAG CCG CTG GGT GTC GTT      809
Phe Pro Gly Ala Glu Ala Arg Val Glu Phe Gln Pro Leu Gly Val Val
    105                 110                 115

GGG GTC ATT AGT CCC TGG AAC TTC CCT ATC GTA CTG GCC TTT GGG CCG      857
Gly Val Ile Ser Pro Trp Asn Phe Pro Ile Val Leu Ala Phe Gly Pro
120                 125                 130                 135

CTG GCC GGC ATA TTC GCA GCA GGT AAT CGC GCC ATG CTC AAG CCG TCC      905
Leu Ala Gly Ile Phe Ala Ala Gly Asn Arg Ala Met Leu Lys Pro Ser
                140                 145                 150

GAG CTT ACC CCG CGG ACT TCT GCC CTG CTT GCG GAG CTA ATT GCT CGT      953
Glu Leu Thr Pro Arg Thr Ser Ala Leu Leu Ala Glu Leu Ile Ala Arg
            155                 160                 165
```

```
TAC TTC GAT GAA ACT GAG CTG ACT ACA GTG CTG GGC GAC GCT GAA GTC        1001
Tyr Phe Asp Glu Thr Glu Leu Thr Thr Val Leu Gly Asp Ala Glu Val
        170                     175                     180

GGT GCG CTG TTC AGT GCT CAG CCT TTC GAT CAT CTG ATC TTC ACC GGC        1049
Gly Ala Leu Phe Ser Ala Gln Pro Phe Asp His Leu Ile Phe Thr Gly
    185                     190                     195

GGC ACT GCC GTG GCC AAG CAC ATC ATG CGT GCC GCG GCG GAT AAC CTA        1097
Gly Thr Ala Val Ala Lys His Ile Met Arg Ala Ala Ala Asp Asn Leu
200                     205                     210                     215

GTG CCC GTT ACC CTG GAA TTG GGT GGC AAA TCG CCG GTG ATC GTT TCC        1145
Val Pro Val Thr Leu Glu Leu Gly Gly Lys Ser Pro Val Ile Val Ser
                220                     225                     230

CGC AGT GCA GAT ATG GCG GAC GTT GCA CAA CGG GTG TTG ACG GTG AAA        1193
Arg Ser Ala Asp Met Ala Asp Val Ala Gln Arg Val Leu Thr Val Lys
            235                     240                     245

ACC TTC AAT GCC GGG CAA ATC TGT CTG GCA CC GTG GGT GAG TCG AAC         1240
Thr Phe Asn Ala Gly Gln Ile Cys Leu Ala    Val Gly Glu Ser Asn
        250                     255     257    454 455

TTG GCG ATG CGC GCA CCC TAC GGA GAA GCG ATC CAC GGA CTG CTC TCT        1288
Leu Ala Met Arg Ala Pro Tyr Gly Glu Ala Ile His Gly Leu Leu Ser
    460                     465                     470

GTC CTC CTT TCA ACG GAG TGT TAG AACCGTTGGT AGTGGTTTTG GACGGGCCCA       1342
Val Leu Leu Ser Thr Glu Cys
475                     480 481

GGAGCATGCG CTTCTGGGCC CGTTTCTTGA GTATTCATTG GATAGTCACG CGTGGTAGCT      1402

TCGAGCCTGC ACAGCTGATG AGCACCCTGG AAGGCGCGCT GTACGCGGAC GACTGGGTTC      1462

ATCTTCGCCA TTCATGACGG AACTCCGTTC CCCAGTACCG CGATGACTAT TTTGCCTCTT      1522

CCGATGTCCG ATTCCACGCC GCCTGACGCT AAGCGGGGGC GGGGGCGCCC GCATCCCAGC      1582

CCAGACAGCA ACAAATGAGT AGGCTCTTGG ATGCCGCGGC GGCTGAGATT GGTAACGGCA      1642

ATTTCGTCAA TGTGACGATG GATTCGATTG CCCGTGCTGC CGGCGTCTCA AAAAAAACGC      1702

TGTACGTCTT GGTGGCGAGC AAGGAAGAAC TCATTTCCCG GTTAGTGGCT CGAGACATGT      1762

CCAACCTTGA GGAATTC                                                     1779
```

FIG. 2f:

```
CTGCAGCCGA GCATCGATTG AGCACTTTAC CCAGCTGCGC TGGCTGACCA TTCAGAATGG       60

CCCGCGGCAC TATCCAATCT AAATCGATCT TCGGGCGCCG CGGGCATCAT GCCCGCGGCG      120

CTCGCCTCAT TTCAATCTCT AACTTGATAA AAACAGAGCT GTTCTCCGGT CTTGGTGGAT      180

CAAGGCCAGT CGCGGAGAGT CTCGAAGAGG AGAGTACAGT GAACGCCGAG TCCACATTGC      240

AACCGCAGGC ATCATCATGC TCTGCTCAGC CACGCTACCG CAGTGTGTCG ATTGGTCATC      300

CTCCGGTTGA GGTTACGCAA GACGCTGGAG GTATTGTCCG G ATG CGT TCT CTC GAG      356
                                               Met Arg Ser Leu Glu
                                                 1               5

GCG CTT CTT CCC TTC CCG GGT CGA ATT CTT GAG CGT CTC GAG CAT TGG        404
Ala Leu Leu Pro Phe Pro Gly Arg Ile Leu Glu Arg Leu Glu His Trp
                  10                  15                  20

GCT AAG ACC CGT CCA GAA CAA ACC TGC GTT GCT GCC AGG GCG GCA AAT        452
Ala Lys Thr Arg Pro Glu Gln Thr Cys Val Ala Ala Arg Ala Ala Asn
              25                  30                  35

GGG GAA TGG CGT CGT ATC AGC TAC GCG GAA ATG TTC CAC AAC GTC CGC        500
Gly Glu Trp Arg Arg Ile Ser Tyr Ala Glu Met Phe His Asn Val Arg
          40                  45                  50

GCC ATC GCA CAG AGC TTG CTT CCT TAC GGA CTA TCG GCA GAG CGT CCG        548
Ala Ile Ala Gln Ser Leu Leu Pro Tyr Gly Leu Ser Ala Glu Arg Pro
      55                  60                  65

CTG CTT ATC GTC TCT GGA AAT GAC CTG GAA CAT CTT CAG CTG GCA TTT        596
Leu Leu Ile Val Ser Gly Asn Asp Leu Glu His Leu Gln Leu Ala Phe
  70                  75                  80                  85

GGG GCT ATG TAT GCG GGC ATT CCC TAT TGC CCG GTG TCT CCT GCT TAT        644
Gly Ala Met Tyr Ala Gly Ile Pro Tyr Cys Pro Val Ser Pro Ala Tyr
                  90                  95                 100

TCA CTG CTG TCG CAA GAT TTG GCG AAG CTG CGT CAC ATC GTA GGT CTT        692
Ser Leu Leu Ser Gln Asp Leu Ala Lys Leu Arg His Ile Val Gly Leu
              105                 110                 115

CTG CAA CCG GGA CTG GTC TTT GCT GCC GAT GCA GCA CCT TTC CAG GGG        740
Leu Gln Pro Gly Leu Val Phe Ala Ala Asp Ala Ala Pro Phe Gln
          120                 125                 130     132

ACAGCAAGCG AACCGGAATT GCCAGCTGGG GCGCCCTCTG GTAAGGTTGG GAAGCCCTGC      800

AAAGTAAACT GGATGGCTTT CTTGCCGCCA AGGATCTGAT GGCGCAGGGG ATCAAGATCT      860

GATCAAGAGA CAGGATGAGG ATCGTTTCGC ATG ATT GAA CAA GAT GGA TTG CAC        914
                                     Met Ile Glu Gln Asp Gly Leu His
                                       1               5

GCA GGT TCT CCG GCC GCT TGG GTG GAG AGG CTA TTC GGC TAT GAC TGG        962
Ala Gly Ser Pro Ala Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp
      10                  15                  20
```

```
GCA CAA CAG ACA ATC GGC TGC TCT GAT GCC GCC GTG TTC CGG CTG TCA      1010
Ala Gln Gln Thr Ile Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser
 25                  30                  35                  40

GCG CAG GGG CGC CCG GTT CTT TTT GTC AAG ACC GAC CTG TCC GGT GCC      1058
Ala Gln Gly Arg Pro Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala
                 45                  50                  55

CTG AAT GAA CTG CAG GAC GAG GCA GCG CGG CTA TCG TGG CTG GCC ACG      1106
Leu Asn Glu Leu Gln Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr
             60                  65                  70

ACG GGC GTT CCT TGC GCA GCT GTG CTC GAC GTT GTC ACT GAA GCG GGA      1154
Thr Gly Val Pro Cys Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly
         75                  80                  85

AGG GAC TGG CTG CTA TTG GGC GAA GTG CCG GGG CAG GAT CTC CTG TCA      1202
Arg Asp Trp Leu Leu Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser
     90                  95                  100

TCT CAC CTT GCT CCT GCC GAG AAA GTA TCC ATC ATG GCT GAT GCA ATG      1250
Ser His Leu Ala Pro Ala Glu Lys Val Ser Ile Met Ala Asp Ala Met
105                 110                 115                 120

CGG CGG CTG CAT ACG CTT GAT CCG GCT ACC TGC CCA TTC GAC CAC CAA      1298
Arg Arg Leu His Thr Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln
                125                 130                 135

GCG AAA CAT CGC ATC GAG CGA GCA CGT ACT CGG ATG GAA GCC GGT CTT      1346
Ala Lys His Arg Ile Glu Arg Ala Arg Thr Arg Met Glu Ala Gly Leu
            140                 145                 150

GTC GAT CAG GAT GAT CTG GAC GAA GAG CAT CAG GGG CTC GCG CCA GCC      1394
Val Asp Gln Asp Asp Leu Asp Glu Glu His Gln Gly Leu Ala Pro Ala
        155                 160                 165

GAA CTG TTC GCC AGG CTC AAG GCG CGC ATG CCC GAC GGC GAG GAT CTC      1442
Glu Leu Phe Ala Arg Leu Lys Ala Arg Met Pro Asp Gly Glu Asp Leu
    170                 175                 180

GTC GTG ACC CAT GGC GAT GCC TGC TTG CCG AAT ATC ATG GTG GAA AAT      1490
Val Val Thr His Gly Asp Ala Cys Leu Pro Asn Ile Met Val Glu Asn
185                 190                 195                 200

GGC CGC TTT TCT GGA TTC ATC GAC TGT GGC CGG CTG GGT GTG GCG GAC      1538
Gly Arg Phe Ser Gly Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp
                205                 210                 215

CGC TAT CAG GAC ATA GCG TTG GCT ACC CGT GAT ATT GCT GAA GAG CTT      1586
Arg Tyr Gln Asp Ile Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu
            220                 225                 230

GGC GGC GAA TGG GCT GAC CGC TTC CTC GTG CTT TAC GGT ATC GCC GCT      1634
Gly Gly Glu Trp Ala Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala
        235                 240                 245
```

```
CCC GAT TCG CAG CGC ATC GCC TTC TAT CGC CTT CTT GAC GAG TTC TTC      1682
Pro Asp Ser Gln Arg Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe
    250                 255                 260             264

TGAGCGGGAC TCTGGGGTTC GAAATGACCG ACCAAGCGAC GCCCCT GTT TTG CAA      1737
                                                  Val Leu Gln
                                                  563     565

TGG CGG TCG GCG AAA GTT GAT GCG CTG TAT CGT GGT GAA GAT CAA TCC      1785
Trp Arg Ser Ala Lys Val Asp Ala Leu Tyr Arg Gly Glu Asp Gln Ser
                570                 575                 580

ATG CTG CGT GAC GAG GCC ACA CTG TGA GTTGGTCAGG GGGGGCTTAC           1832
Met Leu Arg Asp Glu Ala Thr Leu
            585             589

TCGGCGTTTT CCGACACTGC GTTGGTTGCG GCAGTGCGCA CCCCCTGGAT TGATTGCGGG    1892

GGTGCCCTGT CGCTGGTGTC GCCTATCGAC TTAGGGGTAA AGGTCGCTCG CGAAGTTCTG    1952

ATGCGTGCGT CGCTTGAACC ACAAATGGTC GATAGCGTAC TCGCAGGCTC TATGGCTCAA    2012

GCAAGCTTTG ATGCTTACCT GCTCCCGCGG CACATTGGCT TGTACAGCGG TGTTCCCAAG    2072

TCGGTTCCGG CCTTGGGGGT GCAGCGCATT TGCGGCACAG GCTTCGAACT GCTTCGGCAG    2132

GCCGGCGAGC AGATTTCCCA AGGCGCTGAT CACGTGCTGT GTGTCGCGGG CTGCAG        2188
```

FIG. 2g:

```
CTGCAGCCGA GCATCGATTG AGCACTTTAC CCAGCTGCGC TGGCTGACCA TTCAGAATGG      60

CCCGCGGCAC TATCCAATCT AAATCGATCT TCGGGCGCCG CGGGCATCAT GCCCGCGGCG     120

CTCGCCTCAT TTCAATCTCT AACTTGATAA AAACAGAGCT GTTCTCCGGT CTTGGTGGAT     180

CAAGGCCAGT CGCGGAGAGT CTCGAAGAGG AGAGTACAGT GAACGCCGAG TCCACATTGC     240

AACCGCAGGC ATCATCATGC TCTGCTCAGC CACGCTACCG CAGTGTGTCG ATTGGTCATC     300

CTCCGGTTGA GGTTACGCAA GACGCTGGAG GTATTGTCCG G ATG CGT TCT CTC GAG     356
                                               Met Arg Ser Leu Glu
                                                 1               5

GCG CTT CTT CCC TTC CCG GGT CGA ATT CTT GAG CGT CTC GAG CAT TGG       404
Ala Leu Leu Pro Phe Pro Gly Arg Ile Leu Glu Arg Leu Glu His Trp
                  10                  15                  20

GCT AAG ACC CGT CCA GAA CAA ACC TGC GTT GCT GCC AGG GCG GCA AAT       452
Ala Lys Thr Arg Pro Glu Gln Thr Cys Val Ala Ala Arg Ala Ala Asn
             25                  30                  35

GGG GAA TGG CGT CGT ATC AGC TAC GCG GAA ATG TTC CAC AAC GTC CGC       500
Gly Glu Trp Arg Arg Ile Ser Tyr Ala Glu Met Phe His Asn Val Arg
         40                  45                  50

GCC ATC GCA CAG AGC TTG CTT CCT TAC GGA CTA TCG GCA GAG CGT CCG       548
Ala Ile Ala Gln Ser Leu Leu Pro Tyr Gly Leu Ser Ala Glu Arg Pro
     55                  60                  65

CTG CTT ATC GTC TCT GGA AAT GAC CTG GAA CAT CTT CAG CTG GCA TTT       596
Leu Leu Ile Val Ser Gly Asn Asp Leu Glu His Leu Gln Leu Ala Phe
 70                  75                  80                  85

GGG GCT ATG TAT GCG GGC ATT CCC TAT TGC CCG GTG TCT CCT GCT TAT       644
Gly Ala Met Tyr Ala Gly Ile Pro Tyr Cys Pro Val Ser Pro Ala Tyr
                 90                  95                 100

TCA CTG CTG TCG CAA GAT TTG GCG AAG CTG CGT CAC ATC GTA GGT CTT       692
Ser Leu Leu Ser Gln Asp Leu Ala Lys Leu Arg His Ile Val Gly Leu
            105                 110                 115

CTG CAA CCG GGA CTG GTC TTT GCT GCC GAT GCA GCA CCT TTC CAG GGG       740
Leu Gln Pro Gly Leu Val Phe Ala Ala Asp Ala Ala Pro Phe Gln
        120                 125                 130     132

GAGAGGCGGT TTGCGTATTG GGCGCATGCA TAAAAACTGT TGTAATTCAT TAAGCATTCT     800

GCCGACATGG AAGCCATCAC AAACGGCATG ATGAACCTGA ATCGCCAGCG GCATCAGCAC     860

CTTGTCGCCT TGCGTATAAT ATTTGCCCAT GGACGCACAC CGTGGAAACG GATGAAGGCA     920

CGAACCCAGT TGACATAAGC CTGTTCGGTT CGTAAACTGT AATGCAAGTA GCGTATGCGC     980

TCACGCAACT GGTCCAGAAC CTTGACCGAA CGCAGCGGTG GTAACGGCGC AGTGGCGGTT    1040

TTCATGGCTT GTTATGACTG TTTTTTTGTA CAGTCTATGC CTCGGGCATC CAAGCAGCAA    1100
```

```
GCGCGTTACG CCGTGGGTCG ATGTTTGATG TTATGGAGCA GCAACG ATG TTA CGC        1155
                                                   Met Leu Arg
                                                    1

AGC AGC AAC GAT GTT ACG CAG CAG GGC AGT CGC CCT AAA ACA AAG TTA       1203
Ser Ser Asn Asp Val Thr Gln Gln Gly Ser Arg Pro Lys Thr Lys Leu
    5                   10                  15

GGT GGC TCA AGT ATG GGC ATC ATT CGC ACA TGT AGG CTC GGC CCT GAC       1251
Gly Gly Ser Ser Met Gly Ile Ile Arg Thr Cys Arg Leu Gly Pro Asp
 20                 25                  30                  35

CAA GTC AAA TCC ATG CGG GCT GCT CTT GAT CTT TTC GGT CGT GAG TTC       1299
Gln Val Lys Ser Met Arg Ala Ala Leu Asp Leu Phe Gly Arg Glu Phe
                40                  45                  50

GGA GAC GTA GCC ACC TAC TCC CAA CAT CAG CCG GAC TCC GAT TAC CTC       1347
Gly Asp Val Ala Thr Tyr Ser Gln His Gln Pro Asp Ser Asp Tyr Leu
            55                  60                  65

GGG AAC TTG CTC CGT AGT AAG ACA TTC ATC GCG CTT GCT GCC TTC GAC       1395
Gly Asn Leu Leu Arg Ser Lys Thr Phe Ile Ala Leu Ala Ala Phe Asp
        70                  75                  80

CAA GAA GCG GTT GTT GGC GCT CTC GCG GCT TAC GTT CTG CCC AGG TTT       1443
Gln Glu Ala Val Val Gly Ala Leu Ala Ala Tyr Val Leu Pro Arg Phe
    85                  90                  95

GAG CAG CCG CGT AGT GAG ATC TAT ATC TAT GAT CTC GCA GTC TCC GGC       1491
Glu Gln Pro Arg Ser Glu Ile Tyr Ile Tyr Asp Leu Ala Val Ser Gly
100                 105                 110                 115

GAG CAC CGG AGG CAG GGC ATT GCC ACC GCG CTC ATC AAT CTC CTC AAG       1539
Glu His Arg Arg Gln Gly Ile Ala Thr Ala Leu Ile Asn Leu Leu Lys
                120                 125                 130

CAT GAG GCC AAC GCG CTT GGT GCT TAT GTG ATC TAC GTG CAA GCA GAT       1587
His Glu Ala Asn Ala Leu Gly Ala Tyr Val Ile Tyr Val Gln Ala Asp
            135                 140                 145

TAC GGT GAC GAT CCC GCA GTG GCT CTC TAT ACA AAG TTG GGC ATA CGG       1635
Tyr Gly Asp Asp Pro Ala Val Ala Leu Tyr Thr Lys Leu Gly Ile Arg
        150                 155                 160

GAA GAA GTG ATG CAC TTT GAT ATC GAC CCA AGT ACC GCC ACC TAA CAA       1683
Glu Glu Val Met His Phe Asp Ile Asp Pro Ser Thr Ala Thr
    165                 170                 175     177

TTCGTTCAAG CCGAGATCGG CTTCCCCT GTT TTG CAA TGG CGG TCG GCG AAA        1735
                               Val Leu Gln Trp Arg Ser Ala Lys
                               563     565                 570

GTT GAT GCG CTG TAT CGT GGT GAA GAT CAA TCC ATG CTG CGT GAC GAG       1783
Val Asp Ala Leu Tyr Arg Gly Glu Asp Gln Ser Met Leu Arg Asp Glu
                575                 580                 585
```

GCC ACA CTG TGA GTTGGTCAGG GGGGGCTTAC TCGGCGTTTT CCGACACTGC 1835
Ala Thr Leu
589

GTTGGTTGCG GCAGTGCGCA CCCCCTGGAT TGATTGCGGG GGTGCCCTGT CGCTGGTGTC 1895

GCCTATCGAC TTAGGGGTAA AGGTCGCTCG CGAAGTTCTG ATGCGTGCGT CGCTTGAACC 1955

ACAAATGGTC GATAGCGTAC TCGCAGGCTC TATGGCTCAA GCAAGCTTTG ATGCTTACCT 2015

GCTCCCGCGG CACATTGGCT TGTACAGCGG TGTTCCCAAG TCGGTTCCGG CCTTGGGGGT 2075

GCAGCGCATT TGCGGCACAG GCTTCGAACT GCTTCGGCAG GCCGGCGAGC AGATTTCCCA 2135

AGGCGCTGAT CACGTGCTGT GTGTCGCGGG CTGCAG 2171

FIG. 2h:

```
CTGCAGCCGA GCATCGATTG AGCACTTTAC CCAGCTGCGC TGGCTGACCA TTCAGAATGG      60

CCCGCGGCAC TATCCAATCT AAATCGATCT TCGGGCGCCG CGGGCATCAT GCCCGCGGCG     120

CTCGCCTCAT TTCAATCTCT AACTTGATAA AAACAGAGCT GTTCTCCGGT CTTGGTGGAT     180

CAAGGCCAGT CGCGGAGAGT CTCGAAGAGG AGAGTACAGT GAACGCCGAG TCCACATTGC     240

AACCGCAGGC ATCATCATGC TCTGCTCAGC CACGCTACCG CAGTGTGTCG ATTGGTCATC     300

CTCCGGTTGA GGTTACGCAA GACGCTGGAG GTATTGTCCG G ATG CGT TCT CTC GAG     356
                                              Met Arg Ser Leu Glu
                                               1               5

GCG CTT CTT CCC TTC CCG GGT CGA ATT CTT GAG CGT CTC GAG CAT TGG       404
Ala Leu Leu Pro Phe Pro Gly Arg Ile Leu Glu Arg Leu Glu His Trp
                 10                  15                  20

GCT AAG ACC CGT CCA GAA CAA ACC TGC GTT GCT GCC AGG GCG GCA AAT       452
Ala Lys Thr Arg Pro Glu Gln Thr Cys Val Ala Ala Arg Ala Ala Asn
             25                  30                  35

GGG GAA TGG CGT CGT ATC AGC TAC GCG GAA ATG TTC CAC AAC GTC CGC       500
Gly Glu Trp Arg Arg Ile Ser Tyr Ala Glu Met Phe His Asn Val Arg
         40                  45                  50

GCC ATC GCA CAG AGC TTG CTT CCT TAC GGA CTA TCG GCA GAG CGT CCG       548
Ala Ile Ala Gln Ser Leu Leu Pro Tyr Gly Leu Ser Ala Glu Arg Pro
     55                  60                  65

CTG CTT ATC GTC TCT GGA AAT GAC CTG GAA CAT CTT CAG CTG GCA TTT       596
Leu Leu Ile Val Ser Gly Asn Asp Leu Glu His Leu Gln Leu Ala Phe
 70                  75                  80                  85

GGG GCT ATG TAT GCG GGC ATT CCC TAT TGC CCG GTG TCT CCT GCT TAT       644
Gly Ala Met Tyr Ala Gly Ile Pro Tyr Cys Pro Val Ser Pro Ala Tyr
                 90                  95                 100

TCA CTG CTG TCG CAA GAT TTG GCG AAG CTG CGT CAC ATC GTA GGT CTT       692
Ser Leu Leu Ser Gln Asp Leu Ala Lys Leu Arg His Ile Val Gly Leu
            105                 110                 115

CTG CAA CCG GGA CTG GTC TTT GCT GCC GAT GCA GCA CCT TTC CAG CGC       740
Leu Gln Pro Gly Leu Val Phe Ala Ala Asp Ala Ala Pro Phe Gln Arg
        120                 125                 130         133

GCT GTT TTG CAA TGG CGG TCG GCG AAA GTT GAT GCG CTG TAT CGT GGT       788
Ala Val Leu Gln Trp Arg Ser Ala Lys Val Asp Ala Leu Tyr Arg Gly
562         565                 570                 575

GAA GAT CAA TCC ATG CTG CGT GAC GAG GCC ACA CTG TGA GTTGGTCAGG        837
Glu Asp Gln Ser Met Leu Arg Asp Glu Ala Thr Leu
        580                 585             589

GGGGGCTTAC TCGGCGTTTT CCGACACTGC GTTGGTTGCG GCAGTGCGCA CCCCCTGGAT     897

TGATTGCGGG GGTGCCCTGT CGCTGGTGTC GCCTATCGAC TTAGGGGTAA AGGTCGCTCG     957
```

```
CGAAGTTCTG ATGCGTGCGT CGCTTGAACC ACAAATGGTC GATAGCGTAC TCGCAGGCTC    1017
TATGGCTCAA GCAAGCTTTG ATGCTTACCT GCTCCCGCGG CACATTGGCT TGTACAGCGG    1077
TGTTCCCAAG TCGGTTCCGG CCTTGGGGGT GCAGCGCATT TGCGGCACAG GCTTCGAACT    1137
GCTTCGGCAG GCCGGCGAGC AGATTTCCCA AGGCGCTGAT CACGTGCTGT GTGTCGCGGG    1197
CTGCAG                                                               1203
```

FIG. 2i:

```
GAATTCCCCT GGCGACGAAA GGGCGGCAGG CCGCATGGCC ACGGCTGGGC GGTAACTGAT      60

GCTTGCGTTA ATCGTTAACC GTTTGAAATT CCTTGCCAAA TTTCGGCGAG AGAATCATGC     120

GGGTACGCCT TTCCGTGCGC TTTGATCTGC GCTTCCGTGC CTTGAATCAG AAAAATAGTT     180

AATTGACAGA ACTATAGGTT CGCAGTAGCT TTTGCTCACC CACCAAATCC ACAGCACTGG     240

GGTGCACG ATG AAT AGC TAC GAT GGC CGT TGG TCT ACC GTT GAT GTG AAG       290
         Met Asn Ser Tyr Asp Gly Arg Trp Ser Thr Val Asp Val Lys
           1               5                  10

GTT GAA GAA GGT ATC GCT TGG GTC ACG CTG AAC CGC CCG GAG AAG CGC       338
Val Glu Glu Gly Ile Ala Trp Val Thr Leu Asn Arg Pro Glu Lys Arg
 15                  20                  25                  30

AAC GCA ATG AGC CCA ACT CTC AAT CGA GAG ATG GTC GAG GTT CTG GAG       386
Asn Ala Met Ser Pro Thr Leu Asn Arg Glu Met Val Glu Val Leu Glu
                 35                  40                  45

GTG CTG GAG CAG GAC GCA GAT GCT CGC GTG CTT GTT CTG ACT GGT GCA       434
Val Leu Glu Gln Asp Ala Asp Ala Arg Val Leu Val Leu Thr Gly Ala
            50                  55                  60

GGC GAA TCC TGG ACC GCG GGC ATG GAC CTG AAG GAG TAT TTC CGC GAG       482
Gly Glu Ser Trp Thr Ala Gly Met Asp Leu Lys Glu Tyr Phe Arg Glu
         65                  70                  75

ACC GAT GCT GGC CCC GAA ATT CTG CAA GAG AAG ATT CGT CGGGGACAGC       531
Thr Asp Ala Gly Pro Glu Ile Leu Gln Glu Lys Ile Arg
     80                  85                  90  91

AAGCGAACCG GAATTGCCAG CTGGGGCGCC CTCTGGTAAG GTTGGGAAGC CCTGCAAAGT     591

AAACTGGATG GCTTTCTTGC CGCCAAGGAT CTGATGGCGC AGGGGATCAA GATCTGATCA     651

AGAGACAGGA TGAGGATCGT TTCGC ATG ATT GAA CAA GAT GGA TTG CAC GCA       703
                           Met Ile Glu Gln Asp Gly Leu His Ala
                             1               5

GGT TCT CCG GCC GCT TGG GTG GAG AGG CTA TTC GGC TAT GAC TGG GCA       751
Gly Ser Pro Ala Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala
 10                  15                  20                  25

CAA CAG ACA ATC GGC TGC TCT GAT GCC GCC GTG TTC CGG CTG TCA GCG       799
Gln Gln Thr Ile Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser Ala
                 30                  35                  40

CAG GGG CGC CCG GTT CTT TTT GTC AAG ACC GAC CTG TCC GGT GCC CTG       847
Gln Gly Arg Pro Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala Leu
            45                  50                  55

AAT GAA CTG CAG GAC GAG GCA GCG CGG CTA TCG TGG CTG GCC ACG ACG       895
Asn Glu Leu Gln Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Thr
         60                  65                  70
```

```
GGC GTT CCT TGC GCA GCT GTG CTC GAC GTT GTC ACT GAA GCG GGA AGG        943
Gly Val Pro Cys Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly Arg
     75                  80                  85

GAC TGG CTG CTA TTG GGC GAA GTG CCG GGG CAG GAT CTC CTG TCA TCT        991
Asp Trp Leu Leu Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser Ser
 90                  95                 100                 105

CAC CTT GCT CCT GCC GAG AAA GTA TCC ATC ATG GCT GAT GCA ATG CGG       1039
His Leu Ala Pro Ala Glu Lys Val Ser Ile Met Ala Asp Ala Met Arg
                110                 115                 120

CGG CTG CAT ACG CTT GAT CCG GCT ACC TGC CCA TTC GAC CAC CAA GCG       1087
Arg Leu His Thr Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln Ala
            125                 130                 135

AAA CAT CGC ATC GAG CGA GCA CGT ACT CGG ATG GAA GCC GGT CTT GTC       1135
Lys His Arg Ile Glu Arg Ala Arg Thr Arg Met Glu Ala Gly Leu Val
        140                 145                 150

GAT CAG GAT GAT CTG GAC GAA GAG CAT CAG GGG CTC GCG CCA GCC GAA       1183
Asp Gln Asp Asp Leu Asp Glu Glu His Gln Gly Leu Ala Pro Ala Glu
    155                 160                 165

CTG TTC GCC AGG CTC AAG GCG CGC ATG CCC GAC GGC GAG GAT CTC GTC       1231
Leu Phe Ala Arg Leu Lys Ala Arg Met Pro Asp Gly Glu Asp Leu Val
170                 175                 180                 185

GTG ACC CAT GGC GAT GCC TGC TTG CCG AAT ATC ATG GTG GAA AAT GGC       1279
Val Thr His Gly Asp Ala Cys Leu Pro Asn Ile Met Val Glu Asn Gly
                190                 195                 200

CGC TTT TCT GGA TTC ATC GAC TGT GGC CGG CTG GGT GTG GCG GAC CGC       1327
Arg Phe Ser Gly Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp Arg
            205                 210                 215

TAT CAG GAC ATA GCG TTG GCT ACC CGT GAT ATT GCT GAA GAG CTT GGC       1375
Tyr Gln Asp Ile Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu Gly
        220                 225                 230

GGC GAA TGG GCT GAC CGC TTC CTC GTG CTT TAC GGT ATC GCC GCT CCC       1423
Gly Glu Trp Ala Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala Pro
    235                 240                 245

GAT TCG CAG CGC ATC GCC TTC TAT CGC CTT CTT GAC GAG TTC TTC TGA       1471
Asp Ser Gln Arg Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe
250                 255                 260                 264

GCGGGACTCT GGGGTTCGAA ATGACCGACC AAGCGACGCC CC GAG CAG GGC ATG        1525
                                               Glu Gln Gly Met
                                               255

AAG CAG TTC CTT GAC GAG AAA AGC ATC AAG CCG GGC TTG CAG ACC TAC       1573
Lys Gln Phe Leu Asp Glu Lys Ser Ile Lys Pro Gly Leu Gln Thr Tyr
    260                 265                 270
```

```
AAG CGC TGA TAAATGCGCC GGGGCCCTCG CTGCGCCCCC GGCCTTCCAA TAATGACAAT    1632
Lys Arg
275 276

AATGAGGAGT GCCCAATGTT TCACGTGCCC CTGCTTATTG GTGGTAAGCC TTGTTCAGCA    1692

TCTGATGAGC GCACCTTCGA GCGTCGTAGC CCGCTGACCG GAGAAGTGGT ATCGCGCGTC    1752

GCTGCTGCCA GTTTGGAAGA TGCGGACGCC GCAGTGGCCG CTGCACAGGC TGCGTTTCCT    1812

GAATGGGCGG CGCTTGCTCC GAGCGAACGC CGTGCCCGAC TGCTGCGAGC GGCGGATCTT    1872

CTAGAGGACC GTTCTTCCGA GTTCACCGCC GCAGCGAGTG AAACTGGCGC AGCGGGAAAC    1932

TGGTATGGGT TTAACGTTTA CCTGGCGGCG GGCATGTTGC GGGGAATTC                1981
```

FIG. 2j:

```
GAATTCCCCT GGCGACGAAA GGGCGGCAGG CCGCATGGCC ACGGCTGGGC GGTAACTGAT      60

GCTTGCGTTA ATCGTTAACC GTTTGAAATT CCTTGCCAAA TTTCGGCGAG AGAATCATGC     120

GGGTACGCCT TTCCGTGCGC TTTGATCTGC GCTTCCGTGC CTTGAATCAG AAAAATAGTT     180

AATTGACAGA ACTATAGGTT CGCAGTAGCT TTTGCTCACC CACCAAATCC ACAGCACTGG     240

GGTGCACG ATG AAT AGC TAC GAT GGC CGT TGG TCT ACC GTT GAT GTG AAG       290
         Met Asn Ser Tyr Asp Gly Arg Trp Ser Thr Val Asp Val Lys
           1               5                  10

GTT GAA GAA GGT ATC GCT TGG GTC ACG CTG AAC CGC CCG GAG AAG CGC       338
Val Glu Glu Gly Ile Ala Trp Val Thr Leu Asn Arg Pro Glu Lys Arg
 15                  20                  25                  30

AAC GCA ATG AGC CCA ACT CTC AAT CGA GAG ATG GTC GAG GTT CTG GAG       386
Asn Ala Met Ser Pro Thr Leu Asn Arg Glu Met Val Glu Val Leu Glu
                35                  40                  45

GTG CTG GAG CAG GAC GCA GAT GCT CGC GTG CTT GTT CTG ACT GGT GCA       434
Val Leu Glu Gln Asp Ala Asp Ala Arg Val Leu Val Leu Thr Gly Ala
            50                  55                  60

GGC GAA TCC TGG ACC GCG GGC ATG GAC CTG AAG GAG TAT TTC CGC GAG       482
Gly Glu Ser Trp Thr Ala Gly Met Asp Leu Lys Glu Tyr Phe Arg Glu
        65                  70                  75

ACC GAT GCT GGC CCC GAA ATT CTG CAA GAG AAG ATT CGT CGGGGGAGAG        531
Thr Asp Ala Gly Pro Glu Ile Leu Gln Glu Lys Ile Arg
    80                  85                  90  91

GCGGTTTGCG TATTGGGCGC ATGCATAAAA ACTGTTGTAA TTCATTAAGC ATTCTGCCGA     591

CATGGAAGCC ATCACAAACG GCATGATGAA CCTGAATCGC CAGCGGCATC AGCACCTTGT     651

CGCCTTGCGT ATAATATTTG CCCATGGACG CACACCGTGG AAACGGATGA AGGCACGAAC     711

CCAGTTGACA TAAGCCTGTT CGGTTCGTAA ACTGTAATGC AAGTAGCGTA TGCGCTCACG     771

CAACTGGTCC AGAACCTTGA CCGAACGCAG CGGTGGTAAC GGCGCAGTGG CGGTTTTCAT     831

GGCTTGTTAT GACTGTTTTT TTGTACAGTC TATGCCTCGG GCATCCAAGC AGCAAGCGCG     891

TTACGCCGTG GGTCGATGTT TGATGTTATG GAGCAGCAAC G ATG TTA CGC AGC AGC      947
                                              Met Leu Arg Ser Ser
                                                1               5

AAC GAT GTT ACG CAG CAG GGC AGT CGC CCT AAA ACA AAG TTA GGT GGC        995
Asn Asp Val Thr Gln Gln Gly Ser Arg Pro Lys Thr Lys Leu Gly Gly
                10                  15                  20

TCA AGT ATG GGC ATC ATT CGC ACA TGT AGG CTC GGC CCT GAC CAA GTC      1043
Ser Ser Met Gly Ile Ile Arg Thr Cys Arg Leu Gly Pro Asp Gln Val
            25                  30                  35
```

```
AAA TCC ATG CGG GCT GCT CTT GAT CTT TTC GGT CGT GAG TTC GGA GAC      1091
Lys Ser Met Arg Ala Ala Leu Asp Leu Phe Gly Arg Glu Phe Gly Asp
         40                  45                  50

GTA GCC ACC TAC TCC CAA CAT CAG CCG GAC TCC GAT TAC CTC GGG AAC      1139
Val Ala Thr Tyr Ser Gln His Gln Pro Asp Ser Asp Tyr Leu Gly Asn
     55                  60                  65

TTG CTC CGT AGT AAG ACA TTC ATC GCG CTT GCT GCC TTC GAC CAA GAA      1187
Leu Leu Arg Ser Lys Thr Phe Ile Ala Leu Ala Ala Phe Asp Gln Glu
 70                  75                  80                  85

GCG GTT GTT GGC GCT CTC GCG GCT TAC GTT CTG CCC AGG TTT GAG CAG      1235
Ala Val Val Gly Ala Leu Ala Ala Tyr Val Leu Pro Arg Phe Glu Gln
                 90                  95                 100

CCG CGT AGT GAG ATC TAT ATC TAT GAT CTC GCA GTC TCC GGC GAG CAC      1283
Pro Arg Ser Glu Ile Tyr Ile Tyr Asp Leu Ala Val Ser Gly Glu His
            105                 110                 115

CGG AGG CAG GGC ATT GCC ACC GCG CTC ATC AAT CTC CTC AAG CAT GAG      1331
Arg Arg Gln Gly Ile Ala Thr Ala Leu Ile Asn Leu Leu Lys His Glu
        120                 125                 130

GCC AAC GCG CTT GGT GCT TAT GTG ATC TAC GTG CAA GCA GAT TAC GGT      1379
Ala Asn Ala Leu Gly Ala Tyr Val Ile Tyr Val Gln Ala Asp Tyr Gly
    135                 140                 145

GAC GAT CCC GCA GTG GCT CTC TAT ACA AAG TTG GGC ATA CGG GAA GAA      1427
Asp Asp Pro Ala Val Ala Leu Tyr Thr Lys Leu Gly Ile Arg Glu Glu
150                 155                 160                 165

GTG ATG CAC TTT GAT ATC GAC CCA AGT ACC GCC ACC TAA CAATTCGTTC       1476
Val Met His Phe Asp Ile Asp Pro Ser Thr Ala Thr
                170                 175     177

AAGCCGAGAT CGGCTTCCCC GAG CAG GGC ATG AAG CAG TTC CTT GAC GAG        1526
                      Glu Gln Gly Met Lys Gln Phe Leu Asp Glu
                      255                 260

AAA AGC ATC AAG CCG GGC TTG CAG ACC TAC AAG CGC TGA TAAATGCGCC       1575
Lys Ser Ile Lys Pro Gly Leu Gln Thr Tyr Lys Arg
265                 270                 275 276

GGGGCCCTCG CTGCGCCCCC GGCCTTCCAA TAATGACAAT AATGAGGAGT GCCCAATGTT    1635

TCACGTGCCC CTGCTTATTG GTGGTAAGCC TTGTTCAGCA TCTGATGAGC GCACCTTCGA    1695

GCGTCGTAGC CCGCTGACCG GAGAAGTGGT ATCGCGCGTC GCTGCTGCCA GTTTGGAAGA    1755

TGCGGACGCC GCAGTGGCCG CTGCACAGGC TGCGTTTCCT GAATGGGCGG CGCTTGCTCC    1815

GAGCGAACGC CGTGCCCGAC TGCTGCGAGC GGCGGATCTT CTAGAGGACC GTTCTTCCGA    1875
```

GTTCACCGCC GCAGCGAGTG AAACTGGCGC AGCGGGAAAC TGGTATGGGT TTAACGTTTA 1935

CCTGGCGGCG GGCATGTTGC GGGGAATTC 1964

FIG. 2k:

```
GAATTCCCCT GGCGACGAAA GGGCGGCAGG CCGCATGGCC ACGGCTGGGC GGTAACTGAT      60

GCTTGCGTTA ATCGTTAACC GTTTGAAATT CCTTGCCAAA TTTCGGCGAG AGAATCATGC     120

GGGTACGCCT TTCCGTGCGC TTTGATCTGC GCTTCCGTGC CTTGAATCAG AAAAATAGTT     180

AATTGACAGA ACTATAGGTT CGCAGTAGCT TTTGCTCACC CACCAAATCC ACAGCACTGG     240

GGTGCACG ATG AAT AGC TAC GAT GGC CGT TGG TCT ACC GTT GAT GTG AAG       290
         Met Asn Ser Tyr Asp Gly Arg Trp Ser Thr Val Asp Val Lys
           1               5                  10

GTT GAA GAA GGT ATC GCT TGG GTC ACG CTG AAC CGC CCG GAG AAG CGC       338
Val Glu Glu Gly Ile Ala Trp Val Thr Leu Asn Arg Pro Glu Lys Arg
 15                  20                  25                  30

AAC GCA ATG AGC CCA ACT CTC AAT CGA GAG ATG GTC GAG GTT CTG GAG       386
Asn Ala Met Ser Pro Thr Leu Asn Arg Glu Met Val Glu Val Leu Glu
                 35                  40                  45

GTG CTG GAG CAG GAC GCA GAT GCT CGC GTG CTT GTT CTG ACT GGT GCA       434
Val Leu Glu Gln Asp Ala Asp Ala Arg Val Leu Val Leu Thr Gly Ala
            50                  55                  60

GGC GAA TCC TGG ACC GCG GGC ATG GAC CTG AAG GAG TAT TTC CGC GAG       482
Gly Glu Ser Trp Thr Ala Gly Met Asp Leu Lys Glu Tyr Phe Arg Glu
         65                  70                  75

ACC GAT GCT GGC CCC GAA ATT CTG CAA GAG AAG ATT CGT CGC GAG CAG       530
Thr Asp Ala Gly Pro Glu Ile Leu Gln Glu Lys Ile Arg Arg Glu Gln
     80                  85                  90      92 255

GGC ATG AAG CAG TTC CTT GAC GAG AAA AGC ATC AAG CCG GGC TTG CAG       578
Gly Met Lys Gln Phe Leu Asp Glu Lys Ser Ile Lys Pro Gly Leu Gln
            260                 265                 270

ACC TAC AAG CGC TGA TAAATGCGCC GGGGCCCTCG CTGCGCCCCC GGCCTTCCAA       633
Thr Tyr Lys Arg
        275 276

TAATGACAAT AATGAGGAGT GCCCAATGTT TCACGTGCCC CTGCTTATTG GTGGTAAGCC     693

TTGTTCAGCA TCTGATGAGC GCACCTTCGA GCGTCGTAGC CCGCTGACCG GAGAAGTGGT     753

ATCGCGCGTC GCTGCTGCCA GTTTGGAAGA TGCGGACGCC GCAGTGGCCG CTGCACAGGC     813

TGCGTTTCCT GAATGGGCGG CGCTTGCTCC GAGCGAACGC CGTGCCCGAC TGCTGCGAGC     873

GGCGGATCTT CTAGAGGACC GTTCTTCCGA GTTCACCGCC GCAGCGAGTG AAACTGGCGC     933

AGCGGGAAAC TGGTATGGGT TTAACGTTTA CCTGGCGGCG GGCATGTTGC GGGGAATTC      992
```

FIG. 21:

```
GAATTCCAAT AATGACAATA ATGAGGAGTG CCCA ATG TTT CAC GTG CCC CTG CTT      55
                                      Met Phe His Val Pro Leu Leu
                                        1               5

ATT GGT GGT AAG CCT TGT TCA GCA TCT GAT GAG CGC ACC TTC GAG CGT      103
Ile Gly Gly Lys Pro Cys Ser Ala Ser Asp Glu Arg Thr Phe Glu Arg
         10                  15                  20

CGT AGC CCG CTG ACC GGA GAA GTG GTA TCG CGC GTC GCT GCT GCC AGT      151
Arg Ser Pro Leu Thr Gly Glu Val Val Ser Arg Val Ala Ala Ala Ser
     25                  30                  35

TTG GAA GAT GCG GAC GCC GCA GTG GCC GCT GCA CAG GCT GCG TTT CCT      199
Leu Glu Asp Ala Asp Ala Ala Val Ala Ala Ala Gln Ala Ala Phe Pro
 40                  45                  50                  55

GAA TGG GCG GCG CTT GCT CCG AGC GAA CGC CGT GCC CGA CTG CTG CGA      247
Glu Trp Ala Ala Leu Ala Pro Ser Glu Arg Arg Ala Arg Leu Leu Arg
                 60                  65                  70

GCG GCG GAT CTT CTA GAG GAC CGT TCT TCC GAG TTC ACC GCC GCA GCG      295
Ala Ala Asp Leu Leu Glu Asp Arg Ser Ser Glu Phe Thr Ala Ala Ala
             75                  80                  85

AGT GAA ACT GGC GCA GCG GGA AAC TGG TAT GGG TTT AAC GTT TAC CTG      343
Ser Glu Thr Gly Ala Ala Gly Asn Trp Tyr Gly Phe Asn Val Tyr Leu
         90                  95                  100

GCG GCG GGC ATG TTG CGG GAA GCC GCG GCC ATG ACC ACA CAG ATT CAG      391
Ala Ala Gly Met Leu Arg Glu Ala Ala Ala Met Thr Thr Gln Ile Gln
     105                 110                 115

GGC GAT GTC ATT CCG TCC AAT GTG CCC GGT AGC TTT GCC ATG GCG GTT      439
Gly Asp Val Ile Pro Ser Asn Val Pro Gly Ser Phe Ala Met Ala Val
120                 125                 130                 135

CGA CAG CCA TGT GGC GTG GTG CTC GGT ATT GCG CCT TGG AAT GCT CCG      487
Arg Gln Pro Cys Gly Val Val Leu Gly Ile Ala Pro Trp Asn Ala Pro
                140                 145                 150

GTA ATC CTT GGC GTA CGG GCT GTT GCG ATG CCG TTG GCA TGC GGC AAT      535
Val Ile Leu Gly Val Arg Ala Val Ala Met Pro Leu Ala Cys Gly Asn
            155                 160                 165

ACC GTG GTG TTG AAA AGC TCT GAG CTG AGT CCC TTT ACC CAT CGC CTG      583
Thr Val Val Leu Lys Ser Ser Glu Leu Ser Pro Phe Thr His Arg Leu
        170                 175                 180

ATT GGT CAG GTG TTG CAT GAT GCT GGT CTG GGG GAT GGC GTG GTG AAT      631
Ile Gly Gln Val Leu His Asp Ala Gly Leu Gly Asp Gly Val Val Asn
    185                 190                 195

GTC ATC AGC AAT GCC CCG CAA GAC GCT CCT GCG GTG GTG GAG CGA CTG      679
Val Ile Ser Asn Ala Pro Gln Asp Ala Pro Ala Val Val Glu Arg Leu
200                 205                 210                 215
```

```
ATT GCA AAT CCT GCG GTA CGT CGA GTG AAC TTC ACC GGT TCG ACC CAC        727
Ile Ala Asn Pro Ala Val Arg Arg Val Asn Phe Thr Gly Ser Thr His
                220                 225                 230

GTT GGA CGG ATC ATT GGT GAG CTG TCT GCG CGT CAT CTG AAG CCT GCT        775
Val Gly Arg Ile Ile Gly Glu Leu Ser Ala Arg His Leu Lys Pro Ala
            235                 240                 245

GTG CTG GAA TTA GGT GGT AAG GCT CCG TTC TTG GTC TTG GAC GAT GCC        823
Val Leu Glu Leu Gly Gly Lys Ala Pro Phe Leu Val Leu Asp Asp Ala
        250                 255                 260

GAC CTC GAT GCG GCG GTC GAA GCG GCG GCC TTT GGT GCC TAC TTC AAT        871
Asp Leu Asp Ala Ala Val Glu Ala Ala Ala Phe Gly Ala Tyr Phe Asn
    265                 270                 275

CAG GGT CAA ATC TGC ATG TCC ACT GAG CGT CTG ATT GTG ACA GCA GTC        919
Gln Gly Gln Ile Cys Met Ser Thr Glu Arg Leu Ile Val Thr Ala Val
280                 285                 290                 295

GCA GAC GCC TTT GTT GAA AAG CTG GCG AGG AAG GTC GCC ACA CTG CGT        967
Ala Asp Ala Phe Val Glu Lys Leu Ala Arg Lys Val Ala Thr Leu Arg
                300                 305                 310

GCT GGC GAT CCT AAT GAT CCG CAA TCG GTC TTG GGT TCG TTG ATT GAT       1015
Ala Gly Asp Pro Asn Asp Pro Gln Ser Val Leu Gly Ser Leu Ile Asp
            315                 320                 325

GCC AAT GCA GGT CAA CGC ATC CAG GTT CTG GTC GAT GAT GCG CTC GGG       1063
Ala Asn Ala Gly Gln Arg Ile Gln Val Leu Val Asp Asp Ala Leu
        330                 335                 340     342

GACAGCAAGC GAACCGGAAT TGCCAGCTGG GGCGCCCTCT GGTAAGGTTG GGAAGCCCTG     1123

CAAAGTAAAC TGGATGGCTT TCTTGCCGCC AAGGATCTGA TGGCGCAGGG GATCAAGATC     1183

TGATCAAGAG ACAGGATGAG GATCGTTTCG C ATG ATT GAA CAA GAT GGA TTG         1235
                                  Met Ile Glu Gln Asp Gly Leu
                                    1               5

CAC GCA GGT TCT CCG GCC GCT TGG GTG GAG AGG CTA TTC GGC TAT GAC       1283
His Ala Gly Ser Pro Ala Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp
         10                  15                  20

TGG GCA CAA CAG ACA ATC GGC TGC TCT GAT GCC GCC GTG TTC CGG CTG       1331
Trp Ala Gln Gln Thr Ile Gly Cys Ser Asp Ala Ala Val Phe Arg Leu
     25                  30                  35

TCA GCG CAG GGG CGC CCG GTT CTT TTT GTC AAG ACC GAC CTG TCC GGT       1379
Ser Ala Gln Gly Arg Pro Val Leu Phe Val Lys Thr Asp Leu Ser Gly
 40                  45                  50                  55

GCC CTG AAT GAA CTG CAG GAC GAG GCA GCG CGG CTA TCG TGG CTG GCC       1427
Ala Leu Asn Glu Leu Gln Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala
                60                  65                  70
```

ACG ACG GGC GTT CCT TGC GCA GCT GTG CTC GAC GTT GTC ACT GAA GCG 1475
Thr Thr Gly Val Pro Cys Ala Ala Val Leu Asp Val Val Thr Glu Ala
75 80 85

GGA AGG GAC TGG CTG CTA TTG GGC GAA GTG CCG GGG CAG GAT CTC CTG 1523
Gly Arg Asp Trp Leu Leu Leu Gly Glu Val Pro Gly Gln Asp Leu Leu
90 95 100

TCA TCT CAC CTT GCT CCT GCC GAG AAA GTA TCC ATC ATG GCT GAT GCA 1571
Ser Ser His Leu Ala Pro Ala Glu Lys Val Ser Ile Met Ala Asp Ala
105 110 115

ATG CGG CGG CTG CAT ACG CTT GAT CCG GCT ACC TGC CCA TTC GAC CAC 1619
Met Arg Arg Leu His Thr Leu Asp Pro Ala Thr Cys Pro Phe Asp His
120 125 130 135

CAA GCG AAA CAT CGC ATC GAG CGA GCA CGT ACT CGG ATG GAA GCC GGT 1667
Gln Ala Lys His Arg Ile Glu Arg Ala Arg Thr Arg Met Glu Ala Gly
140 145 150

CTT GTC GAT CAG GAT GAT CTG GAC GAA GAG CAT CAG GGG CTC GCG CCA 1715
Leu Val Asp Gln Asp Asp Leu Asp Glu Glu His Gln Gly Leu Ala Pro
155 160 165

GCC GAA CTG TTC GCC AGG CTC AAG GCG CGC ATG CCC GAC GGC GAG GAT 1763
Ala Glu Leu Phe Ala Arg Leu Lys Ala Arg Met Pro Asp Gly Glu Asp
170 175 180

CTC GTC GTG ACC CAT GGC GAT GCC TGC TTG CCG AAT ATC ATG GTG GAA 1811
Leu Val Val Thr His Gly Asp Ala Cys Leu Pro Asn Ile Met Val Glu
185 190 195

AAT GGC CGC TTT TCT GGA TTC ATC GAC TGT GGC CGG CTG GGT GTG GCG 1859
Asn Gly Arg Phe Ser Gly Phe Ile Asp Cys Gly Arg Leu Gly Val Ala
200 205 210 215

GAC CGC TAT CAG GAC ATA GCG TTG GCT ACC CGT GAT ATT GCT GAA GAG 1907
Asp Arg Tyr Gln Asp Ile Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu
220 225 230

CTT GGC GGC GAA TGG GCT GAC CGC TTC CTC GTG CTT TAC GGT ATC GCC 1955
Leu Gly Gly Glu Trp Ala Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala
235 240 245

GCT CCC GAT TCG CAG CGC ATC GCC TTC TAT CGC CTT CTT GAC GAG TTC 2003
Ala Pro Asp Ser Gln Arg Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe
250 255 260

TTC TGA GCGGGACTCT GGGGTTCGAA ATGACCGACC AAGCGACGCC CG GCC CAG 2057
Phe Ala Gln
264 421

CGC GTC GAT TCG GGC ATT TGC CAT ATC AAT GGA CCG ACT GTG CAT GAC 2105
Arg Val Asp Ser Gly Ile Cys His Ile Asn Gly Pro Thr Val His Asp
425 430 435

```
GAG GCT CAG ATG CCA TTC GGT GGG GTG AAG TCC AGC GGC TAC GGC AGC       2153
Glu Ala Gln Met Pro Phe Gly Gly Val Lys Ser Ser Gly Tyr Gly Ser
    440                 445                 450

TTC GGC AGT CGA GCA TCG ATT GAG CAC TTT ACC CAG CTG CGC TGG CTG       2201
Phe Gly Ser Arg Ala Ser Ile Glu His Phe Thr Gln Leu Arg Trp Leu
455                 460                 465                 470

ACC ATT CAG AAT GGC CCG CGG CAC TAT CCA ATC TAA ATCGATCTTC            2247
Thr Ile Gln Asn Gly Pro Arg His Tyr Pro Ile
            475                 480 481

GGGCGCCGCG GGCATCATGC CCGCGGCGCT CGCCTCATTT CAATCTCTAA CTTGATAAAA    2307

ACAGAGCTGT TCTCCGGTCT TGGTGGATCA AGGCCAGTCG CGGAGAGTCT CGAAGAGGAG    2367

AGTACAGTGA ACGCCGAGTC CACATTGCAA CCGCAGGCAT CATCATGCTC TGCTCAGCCA    2427

CGCTACCGCA GTGTGTCGAT TGGTCATCCT CCGGTTGAGG TTACGCAAGA CGCTGGAGGT    2487

ATTGTCCGGA TGCGTTCTCT CGAGGCGCTT CTTCCCTTCC CGGGTGGAAT TC             2539
```

FIG. 2m:

```
GAATTCCAAT AATGACAATA ATGAGGAGTG CCCA ATG TTT CAC GTG CCC CTG CTT       55
                                      Met Phe His Val Pro Leu Leu
                                        1               5

ATT GGT GGT AAG CCT TGT TCA GCA TCT GAT GAG CGC ACC TTC GAG CGT       103
Ile Gly Gly Lys Pro Cys Ser Ala Ser Asp Glu Arg Thr Phe Glu Arg
          10                  15                  20

CGT AGC CCG CTG ACC GGA GAA GTG GTA TCG CGC GTC GCT GCT GCC AGT       151
Arg Ser Pro Leu Thr Gly Glu Val Val Ser Arg Val Ala Ala Ala Ser
      25                  30                  35

TTG GAA GAT GCG GAC GCC GCA GTG GCC GCT GCA CAG GCT GCG TTT CCT       199
Leu Glu Asp Ala Asp Ala Ala Val Ala Ala Ala Gln Ala Ala Phe Pro
 40                  45                  50                  55

GAA TGG GCG GCG CTT GCT CCG AGC GAA CGC CGT GCC CGA CTG CTG CGA       247
Glu Trp Ala Ala Leu Ala Pro Ser Glu Arg Arg Ala Arg Leu Leu Arg
                60                  65                  70

GCG GCG GAT CTT CTA GAG GAC CGT TCT TCC GAG TTC ACC GCC GCA GCG       295
Ala Ala Asp Leu Leu Glu Asp Arg Ser Ser Glu Phe Thr Ala Ala Ala
            75                  80                  85

AGT GAA ACT GGC GCA GCG GGA AAC TGG TAT GGG TTT AAC GTT TAC CTG       343
Ser Glu Thr Gly Ala Ala Gly Asn Trp Tyr Gly Phe Asn Val Tyr Leu
          90                  95                  100

GCG GCG GGC ATG TTG CGG GAA GCC GCG GCC ATG ACC ACA CAG ATT CAG       391
Ala Ala Gly Met Leu Arg Glu Ala Ala Ala Met Thr Thr Gln Ile Gln
      105                 110                 115

GGC GAT GTC ATT CCG TCC AAT GTG CCC GGT AGC TTT GCC ATG GCG GTT       439
Gly Asp Val Ile Pro Ser Asn Val Pro Gly Ser Phe Ala Met Ala Val
120                 125                 130                 135

CGA CAG CCA TGT GGC GTG GTG CTC GGT ATT GCG CCT TGG AAT GCT CCG       487
Arg Gln Pro Cys Gly Val Val Leu Gly Ile Ala Pro Trp Asn Ala Pro
                140                 145                 150

GTA ATC CTT GGC GTA CGG GCT GTT GCG ATG CCG TTG GCA TGC GGC AAT       535
Val Ile Leu Gly Val Arg Ala Val Ala Met Pro Leu Ala Cys Gly Asn
            155                 160                 165

ACC GTG GTG TTG AAA AGC TCT GAG CTG AGT CCC TTT ACC CAT CGC CTG       583
Thr Val Val Leu Lys Ser Ser Glu Leu Ser Pro Phe Thr His Arg Leu
        170                 175                 180

ATT GGT CAG GTG TTG CAT GAT GCT GGT CTG GGG GAT GGC GTG GTG AAT       631
Ile Gly Gln Val Leu His Asp Ala Gly Leu Gly Asp Gly Val Val Asn
    185                 190                 195

GTC ATC AGC AAT GCC CCG CAA GAC GCT CCT GCG GTG GTG GAG CGA CTG       679
Val Ile Ser Asn Ala Pro Gln Asp Ala Pro Ala Val Val Glu Arg Leu
200                 205                 210                 215
```

```
ATT GCA AAT CCT GCG GTA CGT CGA GTG AAC TTC ACC GGT TCG ACC CAC        727
Ile Ala Asn Pro Ala Val Arg Arg Val Asn Phe Thr Gly Ser Thr His
                220                 225                 230

GTT GGA CGG ATC ATT GGT GAG CTG TCT GCG CGT CAT CTG AAG CCT GCT        775
Val Gly Arg Ile Ile Gly Glu Leu Ser Ala Arg His Leu Lys Pro Ala
            235                 240                 245

GTG CTG GAA TTA GGT GGT AAG GCT CCG TTC TTG GTC TTG GAC GAT GCC        823
Val Leu Glu Leu Gly Gly Lys Ala Pro Phe Leu Val Leu Asp Asp Ala
        250                 255                 260

GAC CTC GAT GCG GCG GTC GAA GCG GCG GCC TTT GGT GCC TAC TTC AAT        871
Asp Leu Asp Ala Ala Val Glu Ala Ala Ala Phe Gly Ala Tyr Phe Asn
    265                 270                 275

CAG GGT CAA ATC TGC ATG TCC ACT GAG CGT CTG ATT GTG ACA GCA GTC        919
Gln Gly Gln Ile Cys Met Ser Thr Glu Arg Leu Ile Val Thr Ala Val
280                 285                 290                 295

GCA GAC GCC TTT GTT GAA AAG CTG GCG AGG AAG GTC GCC ACA CTG CGT        967
Ala Asp Ala Phe Val Glu Lys Leu Ala Arg Lys Val Ala Thr Leu Arg
                300                 305                 310

GCT GGC GAT CCT AAT GAT CCG CAA TCG GTC TTG GGT TCG TTG ATT GAT       1015
Ala Gly Asp Pro Asn Asp Pro Gln Ser Val Leu Gly Ser Leu Ile Asp
            315                 320                 325

GCC AAT GCA GGT CAA CGC ATC CAG GTGGGGAGAG GCGGTTTGCG TATTGGGCGC      1069
Ala Asn Ala Gly Gln Arg Ile Gln
        330                 335

ATGCATAAAA ACTGTTGTAA TTCATTAAGC ATTCTGCCGA CATGGAAGCC ATCACAAACG     1129

GCATGATGAA CCTGAATCGC CAGCGGCATC AGCACCTTGT CGCCTTGCGT ATAATATTTG     1189

CCCATGGACG CACACCGTGG AAACGGATGA AGGCACGAAC CCAGTTGACA TAAGCCTGTT     1249

CGGTTCGTAA ACTGTAATGC AAGTAGCGTA TGCGCTCACG CAACTGGTCC AGAACCTTGA     1309

CCGAACGCAG CGGTGGTAAC GGCGCAGTGG CGGTTTTCAT GGCTTGTTAT GACTGTTTTT     1369

TTGTACAGTC TATGCCTCGG GCATCCAAGC AGCAAGCGCG TTACGCCGTG GGTCGATGTT     1429

TGATGTTATG GAGCAGCAAC G ATG TTA CGC AGC AGC AAC GAT GTT ACG CAG        1480
                       Met Leu Arg Ser Ser Asn Asp Val Thr Gln
                         1               5                  10

CAG GGC AGT CGC CCT AAA ACA AAG TTA GGT GGC TCA AGT ATG GGC ATC       1528
Gln Gly Ser Arg Pro Lys Thr Lys Leu Gly Gly Ser Ser Met Gly Ile
                15                  20                  25

ATT CGC ACA TGT AGG CTC GGC CCT GAC CAA GTC AAA TCC ATG CGG GCT       1576
Ile Arg Thr Cys Arg Leu Gly Pro Asp Gln Val Lys Ser Met Arg Ala
            30                  35                  40
```

```
GCT CTT GAT CTT TTC GGT CGT GAG TTC GGA GAC GTA GCC ACC TAC TCC       1624
Ala Leu Asp Leu Phe Gly Arg Glu Phe Gly Asp Val Ala Thr Tyr Ser
         45                  50                  55

CAA CAT CAG CCG GAC TCC GAT TAC CTC GGG AAC TTG CTC CGT AGT AAG       1672
Gln His Gln Pro Asp Ser Asp Tyr Leu Gly Asn Leu Leu Arg Ser Lys
     60                  65                  70

ACA TTC ATC GCG CTT GCT GCC TTC GAC CAA GAA GCG GTT GTT GGC GCT       1720
Thr Phe Ile Ala Leu Ala Ala Phe Asp Gln Glu Ala Val Val Gly Ala
 75                  80                  85                  90

CTC GCG GCT TAC GTT CTG CCC AGG TTT GAG CAG CCG CGT AGT GAG ATC       1768
Leu Ala Ala Tyr Val Leu Pro Arg Phe Glu Gln Pro Arg Ser Glu Ile
                 95                 100                 105

TAT ATC TAT GAT CTC GCA GTC TCC GGC GAG CAC CGG AGG CAG GGC ATT       1816
Tyr Ile Tyr Asp Leu Ala Val Ser Gly Glu His Arg Arg Gln Gly Ile
            110                 115                 120

GCC ACC GCG CTC ATC AAT CTC CTC AAG CAT GAG GCC AAC GCG CTT GGT       1864
Ala Thr Ala Leu Ile Asn Leu Leu Lys His Glu Ala Asn Ala Leu Gly
        125                 130                 135

GCT TAT GTG ATC TAC GTG CAA GCA GAT TAC GGT GAC GAT CCC GCA GTG       1912
Ala Tyr Val Ile Tyr Val Gln Ala Asp Tyr Gly Asp Asp Pro Ala Val
    140                 145                 150

GCT CTC TAT ACA AAG TTG GGC ATA CGG GAA GAA GTG ATG CAC TTT GAT       1960
Ala Leu Tyr Thr Lys Leu Gly Ile Arg Glu Glu Val Met His Phe Asp
155                 160                 165                 170

ATC GAC CCA AGT ACC GCC ACC TAA CAATTCGTTC AAGCCGAGAT CGGCTTCCCA      2014
Ile Asp Pro Ser Thr Ala Thr
                175     177

A TTG GCC CAG CGC GTC GAT TCG GGC ATT TGC CAT ATC AAT GGA CCG ACT     2063
  Leu Ala Gln Arg Val Asp Ser Gly Ile Cys His Ile Asn Gly Pro Thr
  420                 425                 430                 435

GTG CAT GAC GAG GCT CAG ATG CCA TTC GGT GGG GTG AAG TCC AGC GGC       2111
Val His Asp Glu Ala Gln Met Pro Phe Gly Gly Val Lys Ser Ser Gly
                440                 445                 450

TAC GGC AGC TTC GGC AGT CGA GCA TCG ATT GAG CAC TTT ACC CAG CTG       2159
Tyr Gly Ser Phe Gly Ser Arg Ala Ser Ile Glu His Phe Thr Gln Leu
            455                 460                 465

CGC TGG CTG ACC ATT CAG AAT GGC CCG CGG CAC TAT CCA ATC TAA           2204
Arg Trp Leu Thr Ile Gln Asn Gly Pro Arg His Tyr Pro Ile
        470                 475                 480 481

ATCGATCTTC GGGCGCCGCG GGCATCATGC CCGCGGCGCT CGCCTCATTT CAATCTCTAA     2264

CTTGATAAAA ACAGAGCTGT TCTCCGGTCT TGGTGGATCA AGGCCAGTCG CGGAGAGTCT     2324
```

CGAAGAGGAG AGTACAGTGA ACGCCGAGTC CACATTGCAA CCGCAGGCAT CATCATGCTC 2384

TGCTCAGCCA CGCTACCGCA GTGTGTCGAT TGGTCATCCT CCGGTTGAGG TTACGCAAGA 2444

CGCTGGAGGT ATTGTCCGGA TGCGTTCTCT CGAGGCGCTT CTTCCCTTCC CGGGTGGAAT 2504

TC 2506

FIG. 2n:

```
GAATTCCAAT AATGACAATA ATGAGGAGTG CCCA ATG TTT CAC GTG CCC CTG CTT      55
                                      Met Phe His Val Pro Leu Leu
                                        1               5

ATT GGT GGT AAG CCT TGT TCA GCA TCT GAT GAG CGC ACC TTC GAG CGT      103
Ile Gly Gly Lys Pro Cys Ser Ala Ser Asp Glu Arg Thr Phe Glu Arg
         10                  15                  20

CGT AGC CCG CTG ACC GGA GAA GTG GTA TCG CGC GTC GCT GCT GCC AGT      151
Arg Ser Pro Leu Thr Gly Glu Val Val Ser Arg Val Ala Ala Ala Ser
     25                  30                  35

TTG GAA GAT GCG GAC GCC GCA GTG GCC GCT GCA CAG GCT GCG TTT CCT      199
Leu Glu Asp Ala Asp Ala Ala Val Ala Ala Ala Gln Ala Ala Phe Pro
 40                  45                  50                  55

GAA TGG GCG GCG CTT GCT CCG AGC GAA CGC CGT GCC CGA CTG CTG CGA      247
Glu Trp Ala Ala Leu Ala Pro Ser Glu Arg Arg Ala Arg Leu Leu Arg
                 60                  65                  70

GCG GCG GAT CTT CTA GAG GAC CGT TCT TCC GAG TTC ACC GCC GCA GCG      295
Ala Ala Asp Leu Leu Glu Asp Arg Ser Ser Glu Phe Thr Ala Ala Ala
             75                  80                  85

AGT GAA ACT GGC GCA GCG GGA AAC TGG TAT GGG TTT AAC GTT TAC CTG      343
Ser Glu Thr Gly Ala Ala Gly Asn Trp Tyr Gly Phe Asn Val Tyr Leu
         90                  95                  100

GCG GCG GGC ATG TTG CGG GAA GCC GCG GCC ATG ACC ACA CAG ATT CAG      391
Ala Ala Gly Met Leu Arg Glu Ala Ala Ala Met Thr Thr Gln Ile Gln
    105                 110                 115

GGC GAT GTC ATT CCG TCC AAT GTG CCC GGT AGC TTT GCC ATG GCG GTT      439
Gly Asp Val Ile Pro Ser Asn Val Pro Gly Ser Phe Ala Met Ala Val
120                 125                 130                 135

CGA CAG CCA TGT GGC GTG GTG CTC GGT ATT GCG CCT TGG AAT GCT CCG      487
Arg Gln Pro Cys Gly Val Val Leu Gly Ile Ala Pro Trp Asn Ala Pro
                140                 145                 150

GTA ATC CTT GGC GTA CGG GCT GTT GCG ATG CCG TTG GCA TGC GGC AAT      535
Val Ile Leu Gly Val Arg Ala Val Ala Met Pro Leu Ala Cys Gly Asn
            155                 160                 165

ACC GTG GTG TTG AAA AGC TCT GAG CTG AGT CCC TTT ACC CAT CGC CTG      583
Thr Val Val Leu Lys Ser Ser Glu Leu Ser Pro Phe Thr His Arg Leu
        170                 175                 180

ATT GGT CAG GTG TTG CAT GAT GCT GGT CTG GGG GAT GGC GTG GTG AAT      631
Ile Gly Gln Val Leu His Asp Ala Gly Leu Gly Asp Gly Val Val Asn
    185                 190                 195

GTC ATC AGC AAT GCC CCG CAA GAC GCT CCT GCG GTG GTG GAG CGA CTG      679
Val Ile Ser Asn Ala Pro Gln Asp Ala Pro Ala Val Val Glu Arg Leu
200                 205                 210                 215
```

```
ATT GCA AAT CCT GCG GTA CGT CGA GTG AAC TTC ACC GGT TCG ACC CAC        727
Ile Ala Asn Pro Ala Val Arg Arg Val Asn Phe Thr Gly Ser Thr His
                220                 225                 230

GTT GGA CGG ATC ATT GGT GAG CTG TCT GCG CGT CAT CTG AAG CCT GCT        775
Val Gly Arg Ile Ile Gly Glu Leu Ser Ala Arg His Leu Lys Pro Ala
            235                 240                 245

GTG CTG GAA TTA GGT GGT AAG GCT CCG TTC TTG GTC TTG GAC GAT GCC        823
Val Leu Glu Leu Gly Gly Lys Ala Pro Phe Leu Val Leu Asp Asp Ala
        250                 255                 260

GAC CTC GAT GCG GCG GTC GAA GCG GCG GCC TTT GGT GCC TAC TTC AAT        871
Asp Leu Asp Ala Ala Val Glu Ala Ala Ala Phe Gly Ala Tyr Phe Asn
    265                 270                 275

CAG GGT CAA ATC TGC ATG TCC ACT GAG CGT CTG ATT GTG ACA GCA GTC        919
Gln Gly Gln Ile Cys Met Ser Thr Glu Arg Leu Ile Val Thr Ala Val
280                 285                 290                 295

GCA GAC GCC TTT GTT GAA AAG CTG GCG AGG AAG GTC GCC ACA CTG CGT        967
Ala Asp Ala Phe Val Glu Lys Leu Ala Arg Lys Val Ala Thr Leu Arg
                300                 305                 310

GCT GGC GAT CCT AAT GAT CCG CAA TCG GTC TTG GGT TCG TTG ATT GAT       1015
Ala Gly Asp Pro Asn Asp Pro Gln Ser Val Leu Gly Ser Leu Ile Asp
            315                 320                 325

GCC AAT GCA GGT CAA CGC ATC CAG GTT CTG GTC GAT GAT GCG CTC GCA       1063
Ala Asn Ala Gly Gln Arg Ile Gln Val Leu Val Asp Asp Ala Leu Ala
        330                 335                 340

AAA GGC GCG CAATGGAA TTG GCC CAG CGC GTC GAT TCG GGC ATT TGC CAT      1113
Lys Gly Ala           Leu Ala Gln Arg Val Asp Ser Gly Ile Cys His
    345 346           420                 425                 430

ATC AAT GGA CCG ACT GTG CAT GAC GAG GCT CAG ATG CCA TTC GGT GGG       1161
Ile Asn Gly Pro Thr Val His Asp Glu Ala Gln Met Pro Phe Gly Gly
                435                 440                 445

GTG AAG TCC AGC GGC TAC GGC AGC TTC GGC AGT CGA GCA TCG ATT GAG       1209
Val Lys Ser Ser Gly Tyr Gly Ser Phe Gly Ser Arg Ala Ser Ile Glu
            450                 455                 460

CAC TTT ACC CAG CTG CGC TGG CTG ACC ATT CAG AAT GGC CCG CGG CAC       1257
His Phe Thr Gln Leu Arg Trp Leu Thr Ile Gln Asn Gly Pro Arg His
        465                 470                 475

TAT CCA ATC TAA ATCGATCTTC GGGCGCCGCG GGCATCATGC CCGCGGCGCT           1309
Tyr Pro Ile
    480 481

CGCCTCATTT CAATCTCTAA CTTGATAAAA ACAGAGCTGT TCTCCGGTCT TGGTGGATCA     1369

AGGCCAGTCG CGGAGAGTCT CGAAGAGGAG AGTACAGTGA ACGCCGAGTC CACATTGCAA     1429
```

CCGCAGGCAT CATCATGCTC TGCTCAGCCA CGCTACCGCA GTGTGTCGAT TGGTCATCCT 1489

CCGGTTGAGG TTACGCAAGA CGCTGGAGGT ATTGTCCGGA TGCGTTCTCT CGAGGCGCTT 1549

CTTCCCTTCC CGGGTGGAAT TC 1571

FIG. 2o:

```
GAATTCCGCG GTCGGCGAAA GTTGATGCGC TGTATCGTGG TGAAGATCAA TCCATGCTGC      60

GTGACGAGGC CACACT GTG AGT TGG TCA GGG GGG GCT TAC TCG GCG TTT TCC      112
                  Met Ser Trp Ser Gly Gly Ala Tyr Ser Ala Phe Ser
                    1               5                  10

GAC ACT GCG TTG GTT GCG GCA GTG CGC ACC CCC TGG ATT GAT TGC GGG       160
Asp Thr Ala Leu Val Ala Ala Val Arg Thr Pro Trp Ile Asp Cys Gly
         15                  20                  25

GGT GCC CTG TCG CTG GTG TCG CCT ATC GAC TTA GGG GTA AAG GTC GCT       208
Gly Ala Leu Ser Leu Val Ser Pro Ile Asp Leu Gly Val Lys Val Ala
     30                  35                  40

CGC GAA GTT CTG ATG CGT GCG TCG CTT GAA CCA CAA ATG GTC GAT AGC       256
Arg Glu Val Leu Met Arg Ala Ser Leu Glu Pro Gln Met Val Asp Ser
 45                  50                  55                  60

GTA CTC GCA GGC TCT ATG GCT CAA GCA AGC TTT GAT GCT TAC CTG CTC       304
Val Leu Ala Gly Ser Met Ala Gln Ala Ser Phe Asp Ala Tyr Leu Leu
                 65                  70                  75

CCG CGG CAC ATT GGC TTG TAC AGC GGT GTT CCC AAG TCG GTT CCG GCC       352
Pro Arg His Ile Gly Leu Tyr Ser Gly Val Pro Lys Ser Val Pro Ala
             80                  85                  90

TTG GGG GTG CAG CGC ATT TGC GGC ACA GGC TTC GAA CTG CTT CGG CAG       400
Leu Gly Val Gln Arg Ile Cys Gly Thr Gly Phe Glu Leu Leu Arg Gln
         95                  100                 105

GCC GGC GAG CAG ATT TCC CAA GGC GCT GAT CAC GTG CTG TGT GTC GCG       448
Ala Gly Glu Gln Ile Ser Gln Gly Ala Asp His Val Leu Cys Val Ala
    110                 115                 120

GCA GAG TCC ATG TCG CGT AAC CCC ATC GCG TCG TAT ACA CAC CGG GGC       496
Ala Glu Ser Met Ser Arg Asn Pro Ile Ala Ser Tyr Thr His Arg Gly
125                 130                 135                 140

GGG TTC CGC CTC GGT GCG CCC GTT GAG TTC AAG GAT TTT TTG TGG GAG       544
Gly Phe Arg Leu Gly Ala Pro Val Glu Phe Lys Asp Phe Leu Trp Glu
                145                 150                 155

GCA TTG TTT GAT CCT GCT CCA GGA CTC GAC ATG ATC GCT ACC GCA GAA       592
Ala Leu Phe Asp Pro Ala Pro Gly Leu Asp Met Ile Ala Thr Ala Glu
            160                 165                 170

AAC CTG GGGACAGCAA GCGAACCGGA ATTGCCAGCT GGGGCGCCCT CTGGTAAGGT       648
Asn Leu
    174

TGGGAAGCCC TGCAAAGTAA ACTGGATGGC TTTCTTGCCG CCAAGGATCT GATGGCGCAG     708

GGGATCAAGA TCTGATCAAG AGACAGGATG AGGATCGTTT CGC ATG ATT GAA CAA       763
                                                Met Ile Glu Gln
                                                  1
```

GAT GGA TTG CAC GCA GGT TCT CCG GCC GCT TGG GTG GAG AGG CTA TTC 811
Asp Gly Leu His Ala Gly Ser Pro Ala Ala Trp Val Glu Arg Leu Phe
5 10 15 20

GGC TAT GAC TGG GCA CAA CAG ACA ATC GGC TGC TCT GAT GCC GCC GTG 859
Gly Tyr Asp Trp Ala Gln Gln Thr Ile Gly Cys Ser Asp Ala Ala Val
25 30 35

TTC CGG CTG TCA GCG CAG GGG CGC CCG GTT CTT TTT GTC AAG ACC GAC 907
Phe Arg Leu Ser Ala Gln Gly Arg Pro Val Leu Phe Val Lys Thr Asp
40 45 50

CTG TCC GGT GCC CTG AAT GAA CTG CAG GAC GAG GCA GCG CGG CTA TCG 955
Leu Ser Gly Ala Leu Asn Glu Leu Gln Asp Glu Ala Ala Arg Leu Ser
55 60 65

TGG CTG GCC ACG ACG GGC GTT CCT TGC GCA GCT GTG CTC GAC GTT GTC 1003
Trp Leu Ala Thr Thr Gly Val Pro Cys Ala Ala Val Leu Asp Val Val
70 75 80

ACT GAA GCG GGA AGG GAC TGG CTG CTA TTG GGC GAA GTG CCG GGG CAG 1051
Thr Glu Ala Gly Arg Asp Trp Leu Leu Leu Gly Glu Val Pro Gly Gln
85 90 95 100

GAT CTC CTG TCA TCT CAC CTT GCT CCT GCC GAG AAA GTA TCC ATC ATG 1099
Asp Leu Leu Ser Ser His Leu Ala Pro Ala Glu Lys Val Ser Ile Met
105 110 115

GCT GAT GCA ATG CGG CGG CTG CAT ACG CTT GAT CCG GCT ACC TGC CCA 1147
Ala Asp Ala Met Arg Arg Leu His Thr Leu Asp Pro Ala Thr Cys Pro
120 125 130

TTC GAC CAC CAA GCG AAA CAT CGC ATC GAG CGA GCA CGT ACT CGG ATG 1195
Phe Asp His Gln Ala Lys His Arg Ile Glu Arg Ala Arg Thr Arg Met
135 140 145

GAA GCC GGT CTT GTC GAT CAG GAT GAT CTG GAC GAA GAG CAT CAG GGG 1243
Glu Ala Gly Leu Val Asp Gln Asp Asp Leu Asp Glu Glu His Gln Gly
150 155 160

CTC GCG CCA GCC GAA CTG TTC GCC AGG CTC AAG GCG CGC ATG CCC GAC 1291
Leu Ala Pro Ala Glu Leu Phe Ala Arg Leu Lys Ala Arg Met Pro Asp
165 170 175 180

GGC GAG GAT CTC GTC GTG ACC CAT GGC GAT GCC TGC TTG CCG AAT ATC 1339
Gly Glu Asp Leu Val Val Thr His Gly Asp Ala Cys Leu Pro Asn Ile
185 190 195

ATG GTG GAA AAT GGC CGC TTT TCT GGA TTC ATC GAC TGT GGC CGG CTG 1387
Met Val Glu Asn Gly Arg Phe Ser Gly Phe Ile Asp Cys Gly Arg Leu
200 205 210

GGT GTG GCG GAC CGC TAT CAG GAC ATA GCG TTG GCT ACC CGT GAT ATT 1435
Gly Val Ala Asp Arg Tyr Gln Asp Ile Ala Leu Ala Thr Arg Asp Ile
215 220 225

```
GCT GAA GAG CTT GGC GGC GAA TGG GCT GAC CGC TTC CTC GTG CTT TAC        1483
Ala Glu Glu Leu Gly Gly Glu Trp Ala Asp Arg Phe Leu Val Leu Tyr
    230                 235                 240

GGT ATC GCC GCT CCC GAT TCG CAG CGC ATC GCC TTC TAT CGC CTT CTT        1531
Gly Ile Ala Ala Pro Asp Ser Gln Arg Ile Ala Phe Tyr Arg Leu Leu
245                 250                 255                 260

GAC GAG TTC TTC TGA GCGGGACTCT GGGGTTCGAA ATGACCGACC AAGCGACGCC        1586
Asp Glu Phe Phe
            264

CA TTG AGG GCG CAA GAG GAG AAA TGG ATT GAC CAA GAG ATC GTG GCT         1633
   Leu Arg Ala Gln Glu Glu Lys Trp Ile Asp Gln Glu Ile Val Ala
   197          200                 205                 210

GTT ACG GAT GAA CAG TTC GAT TTA GAG GGC TAC AAC AGT CGA GCA ATT        1681
Val Thr Asp Glu Gln Phe Asp Leu Glu Gly Tyr Asn Ser Arg Ala Ile
            215                 220                 225

GAA CTG CCT CGG AAG GCA AAA TTG TTG ATC GTG ACA GTC ATC CGC GGC        1729
Glu Leu Pro Arg Lys Ala Lys Leu Leu Ile Val Thr Val Ile Arg Gly
        230                 235                 240

CTA GCA GTC TTT GAA GCC CTT TCC CGA TTG AAG CCT GTT CAT TCT GGC        1777
Leu Ala Val Phe Glu Ala Leu Ser Arg Leu Lys Pro Val His Ser Gly
    245                 250                 255

GGG GTG CAG ACT GCG GGC AAC AGC TGT GCC GTA GTG GAC GGC GCC GCG        1825
Gly Val Gln Thr Ala Gly Asn Ser Cys Ala Val Val Asp Gly Ala Ala
260                 265                 270                 275

GCG GCT TTG GTG GCT CGA GAG TCG TCT GCG ACA CAG CCG GTC TTG GCT        1873
Ala Ala Leu Val Ala Arg Glu Ser Ser Ala Thr Gln Pro Val Leu Ala
                280                 285                 290

AGG ATA CTG GCT ACC TCC GTA GTC GGG ATC GAG CCC GAG CAT ATG GGG        1921
Arg Ile Leu Ala Thr Ser Val Val Gly Ile Glu Pro Glu His Met Gly
            295                 300                 305

CTC GGC CCT GCG CCC GCG ATT CGC CTG CTG CTT GCG CGT AGT GAT CTT        1969
Leu Gly Pro Ala Pro Ala Ile Arg Leu Leu Leu Ala Arg Ser Asp Leu
        310                 315                 320

AGT TTG AGG GAT ATC GAC CTC TTT GAG ATA AAC GAG GCG CAG GCC GCC        2017
Ser Leu Arg Asp Ile Asp Leu Phe Glu Ile Asn Glu Ala Gln Ala Ala
    325                 330                 335

CAA GTT CTA GCG GTA CAG CAT GAA TTG GGT ATT GAG CAC TCA AAA CTT        2065
Gln Val Leu Ala Val Gln His Glu Leu Gly Ile Glu His Ser Lys Leu
340                 345                 350                 355

AAT ATT TGG GGC GGG GCC ATT GCA CTT GGA CAC CCG CTT GCC GCG ACC        2113
Asn Ile Trp Gly Gly Ala Ile Ala Leu Gly His Pro Leu Ala Ala Thr
                360                 365                 370
```

```
GGA TTG CGT CTC TGC ATG ACC CTC GCT CAC CAA TTG CAA GCT AAT AAC      2161
Gly Leu Arg Leu Cys Met Thr Leu Ala His Gln Leu Gln Ala Asn Asn
            375                 380                 385

TTT CGA TAT GGA ATT GCC TCG GCA TGC ATT GGT GGG GGA CAG GGG ATG      2209
Phe Arg Tyr Gly Ile Ala Ser Ala Cys Ile Gly Gly Gly Gln Gly Met
        390                 395                 400

GCG GTT CTT TTA GAG AAT CCC CAC TTC GGT TCG TCC TCT GCA CGA AGT      2257
Ala Val Leu Leu Glu Asn Pro His Phe Gly Ser Ser Ser Ala Arg Ser
    405                 410                 415

TCG ATG ATT AAC AGA GTT GAC CAC TAT CCA CTG AGC TAA CGGGCATCTC       2306
Ser Met Ile Asn Arg Val Asp His Tyr Pro Leu Ser
420                 425                 430 431

CTTTGTTGCT TTGAGGTGGC GCACGAAGGA GGGCTCGAAA ATCTCTGCTA AAAACAAGAA    2366

GAAGGAACAG GGAACATGAT TAGTTTCGCT CGTATGGCAG AAAGTTTAGG AGTCCAGGCT    2426

AAACTTGCCC TTGCCTTCGC ACTCGTATTA TGTGTCGGGC TGATTGTTAC CGGCACGGGT    2486

TTCTACAGTG TACATACCTT GTCAGGGTTG GTGGGAATTC                          2526
```

FIG. 2p:

```
GAATTCCGCG GTCGGCGAAA GTTGATGCGC TGTATCGTGG TGAAGATCAA TCCATGCTGC      60

GTGACGAGGC CACACT GTG AGT TGG TCA GGG GGG GCT TAC TCG GCG TTT TCC     112
                  Met Ser Trp Ser Gly Gly Ala Tyr Ser Ala Phe Ser
                    1               5                  10

GAC ACT GCG TTG GTT GCG GCA GTG CGC ACC CCC TGG ATT GAT TGC GGG      160
Asp Thr Ala Leu Val Ala Ala Val Arg Thr Pro Trp Ile Asp Cys Gly
         15                  20                  25

GGT GCC CTG TCG CTG GTG TCG CCT ATC GAC TTA GGG GTA AAG GTC GCT      208
Gly Ala Leu Ser Leu Val Ser Pro Ile Asp Leu Gly Val Lys Val Ala
     30                  35                  40

CGC GAA GTT CTG ATG CGT GCG TCG CTT GAA CCA CAA ATG GTC GAT AGC      256
Arg Glu Val Leu Met Arg Ala Ser Leu Glu Pro Gln Met Val Asp Ser
 45                  50                  55                  60

GTA CTC GCA GGC TCT ATG GCT CAA GCA AGC TTT GAT GCT TAC CTG CTC      304
Val Leu Ala Gly Ser Met Ala Gln Ala Ser Phe Asp Ala Tyr Leu Leu
                 65                  70                  75

CCG CGG CAC ATT GGC TTG TAC AGC GGT GTT CCC AAG TCG GTT CCG GCC      352
Pro Arg His Ile Gly Leu Tyr Ser Gly Val Pro Lys Ser Val Pro Ala
             80                  85                  90

TTG GGG GTG CAG CGC ATT TGC GGC ACA GGC TTC GAA CTG CTT CGG CAG      400
Leu Gly Val Gln Arg Ile Cys Gly Thr Gly Phe Glu Leu Leu Arg Gln
         95                  100                 105

GCC GGC GAG CAG ATT TCC CAA GGC GCT GAT CAC GTG CTG TGT GTC GCG      448
Ala Gly Glu Gln Ile Ser Gln Gly Ala Asp His Val Leu Cys Val Ala
    110                 115                 120

GCA GAG TCC ATG TCG CGT AAC CCC ATC GCG TCG TAT ACA CAC CGG GGC      496
Ala Glu Ser Met Ser Arg Asn Pro Ile Ala Ser Tyr Thr His Arg Gly
125                 130                 135                 140

GGG TTC CGC CTC GGT GCG CCC GTT GAG TTC AAG GAT TTT TTG TGG GAG      544
Gly Phe Arg Leu Gly Ala Pro Val Glu Phe Lys Asp Phe Leu Trp Glu
                145                 150                 155

GCA TTG TTT GAT CCT GCT CCA GGA CTC GAC ATG ATC GCT ACC GCA GAA      592
Ala Leu Phe Asp Pro Ala Pro Gly Leu Asp Met Ile Ala Thr Ala Glu
            160                 165                 170

AAC CTG GGGGAGAGGC GGTTTGCGTA TTGGGCGCAT GCATAAAAAC TGTTGTAATT       648
Asn Leu
    174

CATTAAGCAT TCTGCCGACA TGGAAGCCAT CACAAACGGC ATGATGAACC TGAATCGCCA     708

GCGGCATCAG CACCTTGTCG CCTTGCGTAT AATATTTGCC CATGGACGCA CACCGTGGAA     768

ACGGATGAAG GCACGAACCC AGTTGACATA AGCCTGTTCG GTTCGTAAAC TGTAATGCAA     828

GTAGCGTATG CGCTCACGCA ACTGGTCCAG AACCTTGACC GAACGCAGCG GTGGTAACGG     888
```

```
CGCAGTGGCG GTTTTCATGG CTTGTTATGA CTGTTTTTTT GTACAGTCTA TGCCTCGGGC      948

ATCCAAGC AGCAAGCGCG TTACGCCGTG GGTCGATGTTTG ATGTTATGGA GCAGCAACG      1007

ATG TTA CGC AGC AGC AAC GAT GTT ACG CAG CAG GGC AGT CGC CCT AAA      1055
Met Leu Arg Ser Ser Asn Asp Val Thr Gln Gln Gly Ser Arg Pro Lys
  1               5                  10                  15

ACA AAG TTA GGT GGC TCA AGT ATG GGC ATC ATT CGC ACA TGT AGG CTC      1103
Thr Lys Leu Gly Gly Ser Ser Met Gly Ile Ile Arg Thr Cys Arg Leu
             20                  25                  30

GGC CCT GAC CAA GTC AAA TCC ATG CGG GCT GCT CTT GAT CTT TTC GGT      1151
Gly Pro Asp Gln Val Lys Ser Met Arg Ala Ala Leu Asp Leu Phe Gly
         35                  40                  45

CGT GAG TTC GGA GAC GTA GCC ACC TAC TCC CAA CAT CAG CCG GAC TCC      1199
Arg Glu Phe Gly Asp Val Ala Thr Tyr Ser Gln His Gln Pro Asp Ser
     50                  55                  60

GAT TAC CTC GGG AAC TTG CTC CGT AGT AAG ACA TTC ATC GCG CTT GCT      1247
Asp Tyr Leu Gly Asn Leu Leu Arg Ser Lys Thr Phe Ile Ala Leu Ala
 65                  70                  75                  80

GCC TTC GAC CAA GAA GCG GTT GTT GGC GCT CTC GCG GCT TAC GTT CTG      1295
Ala Phe Asp Gln Glu Ala Val Val Gly Ala Leu Ala Ala Tyr Val Leu
                 85                  90                  95

CCC AGG TTT GAG CAG CCG CGT AGT GAG ATC TAT ATC TAT GAT CTC GCA      1343
Pro Arg Phe Glu Gln Pro Arg Ser Glu Ile Tyr Ile Tyr Asp Leu Ala
            100                 105                 110

GTC TCC GGC GAG CAC CGG AGG CAG GGC ATT GCC ACC GCG CTC ATC AAT      1391
Val Ser Gly Glu His Arg Arg Gln Gly Ile Ala Thr Ala Leu Ile Asn
        115                 120                 125

CTC CTC AAG CAT GAG GCC AAC GCG CTT GGT GCT TAT GTG ATC TAC GTG      1439
Leu Leu Lys His Glu Ala Asn Ala Leu Gly Ala Tyr Val Ile Tyr Val
    130                 135                 140

CAA GCA GAT TAC GGT GAC GAT CCC GCA GTG GCT CTC TAT ACA AAG TTG      1487
Gln Ala Asp Tyr Gly Asp Asp Pro Ala Val Ala Leu Tyr Thr Lys Leu
145                 150                 155                 160

GGC ATA CGG GAA GAA GTG ATG CAC TTT GAT ATC GAC CCA AGT ACC GCC      1535
Gly Ile Arg Glu Glu Val Met His Phe Asp Ile Asp Pro Ser Thr Ala
                165                 170                 175

ACC TAA CAATTCGTTC AAGCCGAGAT CGGCTTCCCA TTG AGG GCG CAA GAG GAG      1589
Thr                                      Leu Arg Ala Gln Glu Glu
177                                      197         200

AAA TGG ATT GAC CAA GAG ATC GTG GCT GTT ACG GAT GAA CAG TTC GAT      1637
Lys Trp Ile Asp Gln Glu Ile Val Ala Val Thr Asp Glu Gln Phe Asp
        205                 210                 215
```

```
TTA GAG GGC TAC AAC AGT CGA GCA ATT GAA CTG CCT CGG AAG GCA AAA      1685
Leu Glu Gly Tyr Asn Ser Arg Ala Ile Glu Leu Pro Arg Lys Ala Lys
    220                 225                 230

TTG TTG ATC GTG ACA GTC ATC CGC GGC CTA GCA GTC TTT GAA GCC CTT      1733
Leu Leu Ile Val Thr Val Ile Arg Gly Leu Ala Val Phe Glu Ala Leu
235                 240                 245                 250

TCC CGA TTG AAG CCT GTT CAT TCT GGC GGG GTG CAG ACT GCG GGC AAC      1781
Ser Arg Leu Lys Pro Val His Ser Gly Gly Val Gln Thr Ala Gly Asn
                255                 260                 265

AGC TGT GCC GTA GTG GAC GGC GCC GCG GCG GCT TTG GTG GCT CGA GAG      1829
Ser Cys Ala Val Val Asp Gly Ala Ala Ala Ala Leu Val Ala Arg Glu
            270                 275                 280

TCG TCT GCG ACA CAG CCG GTC TTG GCT AGG ATA CTG GCT ACC TCC GTA      1877
Ser Ser Ala Thr Gln Pro Val Leu Ala Arg Ile Leu Ala Thr Ser Val
        285                 290                 295

GTC GGG ATC GAG CCC GAG CAT ATG GGG CTC GGC CCT GCG CCC GCG ATT      1925
Val Gly Ile Glu Pro Glu His Met Gly Leu Gly Pro Ala Pro Ala Ile
    300                 305                 310

CGC CTG CTG CTT GCG CGT AGT GAT CTT AGT TTG AGG GAT ATC GAC CTC      1973
Arg Leu Leu Leu Ala Arg Ser Asp Leu Ser Leu Arg Asp Ile Asp Leu
315                 320                 325                 330

TTT GAG ATA AAC GAG GCG CAG GCC GCC CAA GTT CTA GCG GTA CAG CAT      2021
Phe Glu Ile Asn Glu Ala Gln Ala Ala Gln Val Leu Ala Val Gln His
                335                 340                 345

GAA TTG GGT ATT GAG CAC TCA AAA CTT AAT ATT TGG GGC GGG GCC ATT      2069
Glu Leu Gly Ile Glu His Ser Lys Leu Asn Ile Trp Gly Gly Ala Ile
            350                 355                 360

GCA CTT GGA CAC CCG CTT GCC GCG ACC GGA TTG CGT CTC TGC ATG ACC      2117
Ala Leu Gly His Pro Leu Ala Ala Thr Gly Leu Arg Leu Cys Met Thr
        365                 370                 375

CTC GCT CAC CAA TTG CAA GCT AAT AAC TTT CGA TAT GGA ATT GCC TCG      2165
Leu Ala His Gln Leu Gln Ala Asn Asn Phe Arg Tyr Gly Ile Ala Ser
    380                 385                 390

GCA TGC ATT GGT GGG GGA CAG GGG ATG GCG GTT CTT TTA GAG AAT CCC      2213
Ala Cys Ile Gly Gly Gly Gln Gly Met Ala Val Leu Leu Glu Asn Pro
395                 400                 405                 410

CAC TTC GGT TCG TCC TCT GCA CGA AGT TCG ATG ATT AAC AGA GTT GAC      2261
His Phe Gly Ser Ser Ser Ala Arg Ser Ser Met Ile Asn Arg Val Asp
                415                 420                 425

CAC TAT CCA CTG AGC TAA CGGGCATCTC CTTTGTTGCT TTGAGGTGGC             2309
His Tyr Pro Leu Ser
            430 431
```

GCACGAAGGA GGGCTCGAAA ATCTCTGCTA AAAACAAGAA GAAGGAACAG GGAACATGAT 2369

TAGTTTCGCT CGTATGGCAG AAAGTTTAGG AGTCCAGGCT AAACTTGCCC TTGCCTTCGC 2429

ACTCGTATTA TGTGTCGGGC TGATTGTTAC CGGCACGGGT TTCTACAGTG TACATACCTT 2489

GTCAGGGTTG GTGGGAATTC 2509

FIG. 2q:

```
GAATTCCGCG GTCGGCGAAA GTTGATGCGC TGTATCGTGG TGAAGATCAA TCCATGCTGC        60

GTGACGAGGC CACACT GTG AGT TGG TCA GGG GGG GCT TAC TCG GCG TTT TCC       112
                  Met Ser Trp Ser Gly Gly Ala Tyr Ser Ala Phe Ser
                   1               5                  10

GAC ACT GCG TTG GTT GCG GCA GTG CGC ACC CCC TGG ATT GAT TGC GGG         160
Asp Thr Ala Leu Val Ala Ala Val Arg Thr Pro Trp Ile Asp Cys Gly
         15                  20                  25

GGT GCC CTG TCG CTG GTG TCG CCT ATC GAC TTA GGG GTA AAG GTC GCT         208
Gly Ala Leu Ser Leu Val Ser Pro Ile Asp Leu Gly Val Lys Val Ala
     30                  35                  40

CGC GAA GTT CTG ATG CGT GCG TCG CTT GAA CCA CAA ATG GTC GAT AGC         256
Arg Glu Val Leu Met Arg Ala Ser Leu Glu Pro Gln Met Val Asp Ser
 45                  50                  55                  60

GTA CTC GCA GGC TCT ATG GCT CAA GCA AGC TTT GAT GCT TAC CTG CTC         304
Val Leu Ala Gly Ser Met Ala Gln Ala Ser Phe Asp Ala Tyr Leu Leu
             65                  70                  75

CCG CGG CAC ATT GGC TTG TAC AGC GGT GTT CCC AAG TCG GTT CCG GCC         352
Pro Arg His Ile Gly Leu Tyr Ser Gly Val Pro Lys Ser Val Pro Ala
             80                  85                  90

TTG GGG GTG CAG CGC ATT TGC GGC ACA GGC TTC GAA CTG CTT CGG CAG         400
Leu Gly Val Gln Arg Ile Cys Gly Thr Gly Phe Glu Leu Leu Arg Gln
         95                 100                 105

GCC GGC GAG CAG ATT TCC CAA GGC GCT GAT CAC GTG CTG TGT GTC GCG         448
Ala Gly Glu Gln Ile Ser Gln Gly Ala Asp His Val Leu Cys Val Ala
    110                 115                 120

GCA GAG TCC ATG TCG CGT AAC CCC ATC GCG TCG TAT ACA CAC CGG GGC         496
Ala Glu Ser Met Ser Arg Asn Pro Ile Ala Ser Tyr Thr His Arg Gly
125                 130                 135                 140

GGG TTC CGC CTC GGT GCG CCC GTT GAG TTC AAG GAT TTT TTG TGG GAG         544
Gly Phe Arg Leu Gly Ala Pro Val Glu Phe Lys Asp Phe Leu Trp Glu
            145                 150                 155

GCA TTG TTT GAT CCT GCT CCA GGA CTC GAC ATG ATC GCT ACC GCA GAA         592
Ala Leu Phe Asp Pro Ala Pro Gly Leu Asp Met Ile Ala Thr Ala Glu
            160                 165                 170

AAC CTG GCG CGC A TTG AGG GCG CAA GAG GAG AAA TGG ATT GAC CAA GAG       641
Asn Leu Ala Arg   Leu Arg Ala Gln Glu Glu Lys Trp Ile Asp Gln Glu
        175 176   197         200                 205

ATC GTG GCT GTT ACG GAT GAA CAG TTC GAT TTA GAG GGC TAC AAC AGT         689
Ile Val Ala Val Thr Asp Glu Gln Phe Asp Leu Glu Gly Tyr Asn Ser
    210                 215                 220

CGA GCA ATT GAA CTG CCT CGG AAG GCA AAA TTG TTG ATC GTG ACA GTC         737
Arg Ala Ile Glu Leu Pro Arg Lys Ala Lys Leu Leu Ile Val Thr Val
225                 230                 235                 240
```

```
ATC CGC GGC CTA GCA GTC TTT GAA GCC CTT TCC CGA TTG AAG CCT GTT        785
Ile Arg Gly Leu Ala Val Phe Glu Ala Leu Ser Arg Leu Lys Pro Val
                245                 250                 255

CAT TCT GGC GGG GTG CAG ACT GCG GGC AAC AGC TGT GCC GTA GTG GAC        833
His Ser Gly Gly Val Gln Thr Ala Gly Asn Ser Cys Ala Val Val Asp
            260                 265                 270

GGC GCC GCG GCG GCT TTG GTG GCT CGA GAG TCG TCT GCG ACA CAG CCG        881
Gly Ala Ala Ala Ala Leu Val Ala Arg Glu Ser Ser Ala Thr Gln Pro
        275                 280                 285

GTC TTG GCT AGG ATA CTG GCT ACC TCC GTA GTC GGG ATC GAG CCC GAG        929
Val Leu Ala Arg Ile Leu Ala Thr Ser Val Val Gly Ile Glu Pro Glu
    290                 295                 300

CAT ATG GGG CTC GGC CCT GCG CCC GCG ATT CGC CTG CTG CTT GCG CGT        977
His Met Gly Leu Gly Pro Ala Pro Ala Ile Arg Leu Leu Leu Ala Arg
305                 310                 315                 320

AGT GAT CTT AGT TTG AGG GAT ATC GAC CTC TTT GAG ATA AAC GAG GCG       1025
Ser Asp Leu Ser Leu Arg Asp Ile Asp Leu Phe Glu Ile Asn Glu Ala
                325                 330                 335

CAG GCC GCC CAA GTT CTA GCG GTA CAG CAT GAA TTG GGT ATT GAG CAC       1073
Gln Ala Ala Gln Val Leu Ala Val Gln His Glu Leu Gly Ile Glu His
            340                 345                 350

TCA AAA CTT AAT ATT TGG GGC GGG GCC ATT GCA CTT GGA CAC CCG CTT       1121
Ser Lys Leu Asn Ile Trp Gly Gly Ala Ile Ala Leu Gly His Pro Leu
        355                 360                 365

GCC GCG ACC GGA TTG CGT CTC TGC ATG ACC CTC GCT CAC CAA TTG CAA       1169
Ala Ala Thr Gly Leu Arg Leu Cys Met Thr Leu Ala His Gln Leu Gln
    370                 375                 380

GCT AAT AAC TTT CGA TAT GGA ATT GCC TCG GCA TGC ATT GGT GGG GGA       1217
Ala Asn Asn Phe Arg Tyr Gly Ile Ala Ser Ala Cys Ile Gly Gly Gly
385                 390                 395                 400

CAG GGG ATG GCG GTT CTT TTA GAG AAT CCC CAC TTC GGT TCG TCC TCT       1265
Gln Gly Met Ala Val Leu Leu Glu Asn Pro His Phe Gly Ser Ser Ser
                405                 410                 415

GCA CGA AGT TCG ATG ATT AAC AGA GTT GAC CAC TAT CCA CTG AGC TAA       1313
Ala Arg Ser Ser Met Ile Asn Arg Val Asp His Tyr Pro Leu Ser
            420                 425                 430 431

CGGGCATCTC CTTTGTTGCT TTGAGGTGGC GCACGAAGGA GGGCTCGAAA ATCTCTGCTA     1373

AAAACAAGAA GAAGGAACAG GGAACATGAT TAGTTTCGCT CGTATGGCAG AAAGTTTAGG     1433

AGTCCAGGCT AAACTTGCCC TTGCCTTCGC ACTCGTATTA TGTGTCGGGC TGATTGTTAC     1493

CGGCACGGGT TTCTACAGTG TACATACCTT GTCAGGGTTG GTGGGAATTC                1543
```

FIG. 2r:

Sequence 1

CTGCAGCCAG GGCTGAAAAG GAGGGATTCA GTGAGGTCAT GAAGGGAGGG GACGGCGCCT 60
GGCTCCAATT GCTCGATGGC GCCGCGATTG AGTGTCTTGG GCGCGGTCTT GGAGAGTTCG 120
GCTAGGGAGA TAAATTTGCT GGCCATGGTG GCGGCCCCTG ATGGGTTGGA TGATTTTCTG 180
CATTCTGCAT CATGAAATTC ATGAAATCAT CACTTTTCGG GGGGTGGGTG CACGGGATTG 240
AAGGTTGCTA GGAGAGTGCA TTGCTCGTAA GCCCAGGAAG CACGCGGGTT TCAGGATGGT 300
GCATGGAAAT GGCATGAGCT TTGCTGGATA TGATTAGAGA CATTAACTAT TTTGGCGGAA 360
TGGAAGCACG ATTCCTCGCC CGGTAGAGCG GTAACCGCGA CATTCAGGAC CGTAAAAAGG 420
AAAGAGCATG CAACTGACCA ACAAGAAAAT CGTCGTCACC GGAGTGTCCT CCGGTATCGG 480
TGCCGAAACT GCCCGCGTTC TGCGCTCTCA CGGCGCCACA GTGATTGGCG TAGATCGCAA 540
CATGCCGAGC CTGACTCTGG ATGCTTTCGT TCAGGCTGAC CTGAGCCATC CTGAAGGCAT 600
CGATAAGGCC ATCGGGACAG CAAGCGAACC GGAATTGCCA GCTGGGGCGC CCTCTGGTAA 660
GGTTGGGAAG CCCTGCAAAG TAAACTGGAT GGCTTTCTTG CCGCCAAGGA TCTGATGGCG 720
CAGGGGATCA AGATCTGATC AAGAGACAGG ATGAGGATCG TTTCGCATGA TTGAACAAGA 780
TGGATTGCAC GCAGGTTCTC CGGCCGCTTG GGTGGAGAGG CTATTCGGCT ATGACTGGGC 840
ACAACAGACA ATCGGCTGCT CTGATGCCGC CGTGTTCCGG CTGTCAGCGC AGGGGCGCCC 900
GGTTCTTTTT GTCAAGACCG ACCTGTCCGG TGCCCTGAAT GAACTGCAGG ACGAGGCAGC 960
GCGGCTATCG TGGCTGGCCA CGACGGGCGT TCCTTGCGCA GCTGTGCTCG ACGTTGTCAC 1020
TGAAGCGGGA AGGGACTGGC TGCTATTGGG CGAAGTGCCG GGGCAGGATC TCCTGTCATC 1080
TCACCTTGCT CCTGCCGAGA AAGTATCCAT CATGGCTGAT GCAATGCGGC GGCTGCATAC 1140
GCTTGATCCG GCTACCTGCC CATTCGACCA CCAAGCGAAA CATCGCATCG AGCGAGCACG 1200
TACTCGGATG GAAGCCGGTC TTGTCGATCA GGATGATCTG GACGAAGAGC ATCAGGGGCT 1260
CGCGCCAGCC GAACTGTTCG CCAGGCTCAA GGCGCGCATG CCCGACGGCG AGGATCTCGT 1320
CGTGACCCAT GGCGATGCCT GCTTGCCGAA TATCATGGTG GAAAATGGCC GCTTTTCTGG 1380
ATTCATCGAC TGTGGCCGGC TGGGTGTGGC GGACCGCTAT CAGGACATAG CGTTGGCTAC 1440
CCGTGATATT GCTGAAGAGC TTGGCGGCGA ATGGGCTGAC CGCTTCCTCG TGCTTTACGG 1500
TATCGCCGCT CCCGATTCGC AGCGCATCGC CTTCTATCGC CTTCTTGACG AGTTCTTCTG 1560
AGCGGGACTC TGGGGTTCGA AATGACCGAC CAAGCGACGC CCTGGCCGCG GTGATTGCAT 1620
TCATGTGTGC TGAGGAGTCA CGTTGGATCA ACGGCATAAA TATTCCAGTG GACGGAGGTT 1680
TGGCATCGAC CTACGTGTAA GTTCGTGGAC GCCCTTTGCA CGCGCACTAT ATCTCTATGC 1740
AGCAGCTGAA AGCAGCTTTG GTTTTGATCG GAGGTAGCGG GCGGAAAGGT GCAGAATGTC 1800
TAAATAATAA AGGATTCTTG TGAAGCTTTA GTTGTCCGTA AACGAAAATA AAAATAAAGA 1860
GGAATGATAT GAAAGCAAGT AGATCAGTCT GCACTTTCAA AATAGCTACC CTGGCAGGCG 1920
CCATTTATGC AGCGCTGCCA ATGTCAGCTG CAAACTCGAT GCAGCTGGAT GTAGGTAGCT 1980
CGGATTGGAC GGTGCGTTGG GGACAACACC CTCAAGTATA GCCTTGCCTC TCGCCTGAAT 2040
GAGCAAGACT CAAGTCTGAC AAATGCGCCG ACTGTCAATG GTTATATCCG GATATTCAAA 2100
GTCAGGGTGA TCGTAACTTT GACCGGGGGC TTGGTATCCA ATCGTCTCGA TATTCTGGCT 2160
GCAG 2164

Sequence 2

```
CTGCAGCCAG GGCTGAAAAG GAGGGATTCA GTGAGGTCAT GAAGGGAGGG GACGGCGCCT   60
GGCTCCAATT GCTCGATGGC GCCGCGATTG AGTGTCTTGG GCGCGGTCTT GGAGAGTTCG  120
GCTAGGGAGA TAAATTTGCT GGCCATGGTG GCGGCCCCTG ATGGGTTGGA TGATTTTCTG  180
CATTCTGCAT CATGAAATTC ATGAAATCAT CACTTTTCGG GGGGTGGGTG CACGGGATTG  240
AAGGTTGCTA GGAGAGTGCA TTGCTCGTAA GCCCAGGAAG CACGCGGGTT TCAGGATGGT  300
GCATGGAAAT GGCATGAGCT TTGCTGGATA TGATTAGAGA CATTAACTAT TTTGGCGGAA  360
TGGAAGCACG ATTCCTCGCC CGGTAGAGCG GTAACCGCGA CATTCAGGAC CGTAAAAAGG  420
AAAGAGCATG CAACTGACCA ACAAGAAAAT CGTCGTCACC GGAGTGTCCT CCGGTATCGG  480
TGCCGAAACT GCCCGCGTTC TGCGCTCTCA CGGCGCCACA GTGATTGGCG TAGATCGCAA  540
CATGCCGAGC CTGACTCTGG ATGCTTTCGT TCAGGCTGAC CTGAGCCATC CTGAGGGGAG  600
AGGCGGTTTG CGTATTGGGC GCATGCATAA AAACTGTTGT AATTCATTAA GCATTCTGCC  660
GACATGGAAG CCATCACAAA CGGCATGATG AACCTGAATC GCCAGCGGCA TCAGCACCTT  720
GTCGCCTTGC GTATAATATT TGCCCATGGA CGCACACCGT GGAAACGGAT GAAGGCACGA  780
ACCCAGTTGA CATAAGCCTG TTCGGTTCGT AAACTGTAAT GCAAGTAGCG TATGCGCTCA  840
CGCAACTGGT CCAGAACCTT GACCGAACGC AGCGGTGGTA ACGGCGCAGT GGCGGTTTTC  900
ATGGCTTGTT ATGACTGTTT TTTTGTACAG TCTATGCCTC GGGCATCCAA GCAGCAAGCG  960
CGTTACGCCG TGGGTCGATG TTTGATGTTA TGGAGCAGCA ACGATGTTAC GCAGCAGCAA 1020
CGATGTTACG CAGCAGGGCA GTCGCCCTAA AACAAAGTTA GGTGGCTCAA GTATGGGCAT 1080
CATTCGCACA TGTAGGCTCG GCCCTGACCA AGTCAAATCC ATGCGGGCTG CTCTTGATCT 1140
TTTCGGTCGT GAGTTCGGAG ACGTAGCCAC CTACTCCCAA CATCAGCCGG ACTCCGATTA 1200
CCTCGGGAAC TTGCTCCGTA GTAAGACATT CATCGCGCTT GCTGCCTTCG ACCAAGAAGC 1260
GGTTGTTGGC GCTCTCGCGG CTTACGTTCT GCCCAGGTTT GAGCAGCCGC GTAGTGAGAT 1320
CTATATCTAT GATCTCGCAG TCTCCGGCGA GCACCGGAGG CAGGGCATTG CCACCGCGCT 1380
CATCAATCTC CTCAAGCATG AGGCCAACGC GCTTGGTGCT TATGTGATCT ACGTGCAAGC 1440
AGATTACGGT GACGATCCCG CAGTGGCTCT CTATACAAAG TTGGGCATAC GGGAAGAAGT 1500
GATGCACTTT GATATCGACC CAAGTACCGC CACCTAACAA TTCGTTCAAG CCGAGATCGG 1560
CTTCCCTGAT TGCATTCATG TGTGCTGAGG AGTCACGTTG GATCAACGGC ATAAATATTC 1620
CAGTGGACGG AGGTTTGGCA TCGACCTACG TGTAAGTTCG TGGACGCCCT TTGCACGCGC 1680
ACTATATCTC TATGCAGCAG CTGAAAGCAG CTTTGGTTTT GATCGGAGGT AGCGGGCGGA 1740
AAGGTGCAGA ATGTCTAAAT AATAAAGGAT TCTTGTGAAG CTTTAGTTGT CCGTAAACGA 1800
AAATAAAAAT AAAGAGGAAT GATATGAAAG CAAGTAGATC AGTCTGCACT TTCAAAATAG 1860
CTACCCTGGC AGGCGCCATT TATGCAGCGC TGCCAATGTC AGCTGCAAAC TCGATGCAGC 1920
TGGATGTAGG TAGCTCGGAT TGGACGGTGC GTTGGGGACA ACACCCTCAA GTATAGCCTT 1980
GCCTCTCGCC TGAATGAGCA AGACTCAAGT CTGACAAATG CGCCGACTGT CAATGGTTAT 2040
ATCCGGATAT TCAAAGTCAG GGTGATCGTA ACTTTGACCG GGGGCTTGGT ATCCAATCGT 2100
CTCGATATTC TGGCTGCAG                                             2119
```

Sequence 3

```
CTGCAGCCAG GGCTGAAAAG GAGGGATTCA GTGAGGTCAT GAAGGGAGGG GACGGCGCCT    60
GGCTCCAATT GCTCGATGGC GCCGCGATTG AGTGTCTTGG GCGCGGTCTT GGAGAGTTCG   120
GCTAGGGAGA TAAATTTGCT GGCCATGGTG GCGGCCCCTG ATGGGTTGGA TGATTTTCTG   180
CATTCTGCAT CATGAAATTC ATGAAATCAT CACTTTTCGG GGGGTGGGTG CACGGGATTG   240
AAGGTTGCTA GGAGAGTGCA TTGCTCGTAA GCCCAGGAAG CACGCGGGTT TCAGGATGGT   300
GCATGGAAAT GGCATGAGCT TTGCTGGATA TGATTAGAGA CATTAACTAT TTTGGCGGAA   360
TGGAAGCACG ATTCCTCGCC CGGTAGAGCG GTAACCGCGA CATTCAGGAC CGTAAAAAGG   420
AAAGAGCATG CAACTGACCA ACAAGAAAAT CGTCGTCACC GGAGTGTCCT CCGGTATCGG   480
TGCCGAAACT GCCCGCGTTC TGCGCTCTCA CGGCGCCACA GTGATTGGCG TAGATCGCAA   540
CATGCCGAGC CTGACTCTGG ATGCTTTCGT TCAGGCTGAC CTGAGCCATC CTGAAGGCAT   600
CGATCAACGG CATAAATATT CCAGTGGACG GAGGTTTGGC ATCGACCTAC GTGTAAGTTC   660
GTGGACGCCC TTTGCACGCG CACTATATCT CTATGCAGCA GCTGAAAGCA GCTTTGGTTT   720
TGATCGGAGG TAGCGGGCGG AAAGGTGCAG AATGTCTAAA TAATAAAGGA TTCTTGTGAA   780
GCTTTAGTTG TCCGTAAACG AAAATAAAAA TAAAGAGGAA TGATATGAAA GCAAGTAGAT   840
CAGTCTGCAC TTTCAAAATA GCTACCCTGG CAGGCGCCAT TTATGCAGCG CTGCCAATGT   900
CAGCTGCAAA CTCGATGCAG CTGGATGTAG GTAGCTCGGA TTGGACGGTG CGTTGGGGAC   960
AACACCCTCA AGTATAGCCT TGCCTCTCGC CTGAATGAGC AAGACTCAAG TCTGACAAAT  1020
GCGCCGACTG TCAATGGTTA TATCCGGATA TTCAAAGTCA GGGTGATCGT AACTTTGACC  1080
GGGGGCTTGG TATCCAATCG TCTCGATATT CTGGCTGCAG                        1120
```

Sequence 4

```
GAATTCCGCG TATCGCCCGG TTCTATCAGC GGGCCGCTTT CGAAAGTCAT GGTGTTAGCC   60
GGTAGGGTCT TTTTCTTGGC CATGCTTGTT GCCTGAACCT TCGTTGACAT AGGGCAGAGG  120
TGCGTTTGCC GCTTCGCTTC GCGATGAACC GCATCGAGAT GCTGAGGTCA GGATTTTTCC  180
TTAACTCGCG TAAGCATTCT GTCATTTTTT TGGTGGCTTT GAACAGCCTG ATGAAAGGTG  240
GTCTCGCCCT TTGAGGCCGA TTCTTGGGCG CTTGGCGGCG TCGAAGCGAT GCTCCACTAC  300
CGATTAAGAT AATTAAAATA AGGAAACCGC ATGGTTTCTT ATGTGAATTT GTCTGGCATA  360
CTCCAGCTCA AGGGCAATTT TTGGGCTATT GGCTGAGCAG TTGCCTCTAT ATGGTTATTC  420
AGAATAACAA TTGACTCCTC AGGAGGTCAG CGATGAGCAT TCTTGGTTTG AATGGTGCCC  480
CGGTCGGAGC TGAGCAGCTG GGCTCGGCTC TTGATCGCAT GAAGAAGGCG CACCTGGAGC  540
AGGGGCCTGC AAACTTGGAG CTGCGTCTGA GTAGGCTGGA TCGTGCGATT GCAATGCTTC  600
TGGAAAATCG TGAAGCAATT GCCGACGCGG TTTCTGCTGA CTTTGGCAAT CGCAGCCGTG  660
AGCAAACACT GCTTTGCGAC ATTGCTGGCT CGGTGGCAAG CCTGAAGGAT AGCCGCGAGC  720
ACGTGGCCAA ATGGATGGAG CCCGAACATC ACAAGGCGAT GTTTCCAGGG GCGGAGGCAC  780
GCGTTGAGTT TCAGCCGCTG GGTGTCGTTG GGGTCATTAG TCCCTGGAAC TTCCCTATCG  840
TACTGGCCTT TGGGCCGCTG GCCGGCATAT TCGCAGCAGG TAATCGCGCC ATGCTCAAGC  900
CGTCCGAGCT TACCCCGCGG ACTTCTGCCC TGCTTGCGGA GCTAATTGCT CGTTACTTCG  960
ATGAAACTGA GCTGACTACA GTGCTGGGCG ACGCTGAAGT CGGTGCGCTG TTCAGTGCTC 1020
AGCCTTTCGA TCATCTGATC TTCACCGGCG GCACTGCCGT GGCCAAGCAC ATCATGCGTG 1080
CCGCGGCGGA TAACCTAGTG CCCGTTACCC TGGAATTGGG TGGCAAATCG CCGGTGATCG 1140
TTTCCCGCAG TGCAGATATG GCGGACGTTG CACAACGGGT GTTGACGGTG AAAACCTTCA 1200
ATGCCGGGCA AATCTGTCTG GCACCGGACT ATGTGCTGCT GCCGGAAGGG ACAGCAAGCG 1260
AACCGGAATT GCCAGCTGGG GCGCCCTCTG GTAAGGTTGG GAAGCCCTGC AAAGTAAACT 1320
GGATGGCTTT CTTGCCGCCA AGGATCTGAT GGCGCAGGGG ATCAAGATCT GATCAAGAGA 1380
CAGGATGAGG ATCGTTTCGC ATGATTGAAC AAGATGGATT GCACGCAGGT TCTCCGGCCG 1440
CTTGGGTGGA GAGGCTATTC GGCTATGACT GGGCACAACA GACAATCGGC TGCTCTGATG 1500
CCGCCGTGTT CCGGCTGTCA GCGCAGGGGC GCCCGGTTCT TTTTGTCAAG ACCGACCTGT 1560
CCGGTGCCCT GAATGAACTG CAGGACGAGG CAGCGCGGCT ATCGTGGCTG GCCACGACGG 1620
GCGTTCCTTG CGCAGCTGTG CTCGACGTTG TCACTGAAGC GGGAAGGGAC TGGCTGCTAT 1680
TGGGCGAAGT GCCGGGGCAG GATCTCCTGT CATCTCACCT TGCTCCTGCC GAGAAAGTAT 1740
CCATCATGGC TGATGCAATG CGGCGGCTGC ATACGCTTGA TCCGGCTACC TGCCCATTCG 1800
ACCACCAAGC GAAACATCGC ATCGAGCGAG CACGTACTCG GATGGAAGCC GGTCTTGTCG 1860
ATCAGGATGA TCTGGACGAA GAGCATCAGG GGCTCGCGCC AGCCGAACTG TTCGCCAGGC 1920
TCAAGGCGCG CATGCCCGAC GGCGAGGATC TCGTCGTGAC CCATGGCGAT GCCTGCTTGC 1980
CGAATATCAT GGTGGAAAAT GGCCGCTTTT CTGGATTCAT CGACTGTGGC CGGCTGGGTG 2040
TGGCGGACCG CTATCAGGAC ATAGCGTTGG CTACCCGTGA TATTGCTGAA GAGCTTGGCG 2100
GCGAATGGGC TGACCGCTTC CTCGTGCTTT ACGGTATCGC CGCTCCCGAT TCGCAGCGCA 2160
```

TCGCCTTCTA TCGCCTTCTT GACGAGTTCT TCTGAGCGGG ACTCTGGGGT TCGAAATGAC 2220
CGACCAAGCG ACGCCCGCCA TGCCAAGCCT GTTCTCGTGC AAAGTCCTGT GGGTGAGTCG 2280
AACTTGGCGA TGCGCGCACC CTACGGAGAA GCGATCCACG GACTGCTCTC TGTCCTCCTT 2340
TCAACGGAGT GTTAGAACCG TTGGTAGTGG TTTTGGACGG GCCCAGGAGC ATGCGCTTCT 2400
GGGCCCGTTT CTTGAGTATT CATTGGATAG TCACGCGTGG TAGCTTCGAG CCTGCACAGC 2460
TGATGAGCAC CCTGGAAGGC GCGCTGTACG CGGACGACTG GGTTCATCTT CGCCATTCAT 2520
GACGGAACTC CGTTCCCCAG TACCGCGATG ACTATTTTGC CTCTTCCGAT GTCCGATTCC 2580
ACGCCGCCTG ACGCTAAGCG GGGGCGGGGG CGCCCGCATC CCAGCCCAGA CAGCAACAAA 2640
TGAGTAGGCT CTTGGATGCC GCGGCGGCTG AGATTGGTAA CGGCAATTTC GTCAATGTGA 2700
CGATGGATTC GATTGCCCGT GCTGCCGGCG TCTCAAAAAA AACGCTGTAC GTCTTGGTGG 2760
CGAGCAAGGA AGAACTCATT TCCCGGTTAG TGGCTCGAGA CATGTCCAAC CTTGAGGAAT 2820
TC 2822

Sequence 5

GAATTCCGCG TATCGCCCGG TTCTATCAGC GGGCCGCTTT CGAAAGTCAT GGTGTTAGCC 60
GGTAGGGTCT TTTTCTTGGC CATGCTTGTT GCCTGAACCT TCGTTGACAT AGGGCAGAGG 120
TGCGTTTGCC GCTTCGCTTC GCGATGAACC GCATCGAGAT GCTGAGGTCA GGATTTTTCC 180
TTAACTCGCG TAAGCATTCT GTCATTTTTT TGGTGGCTTT GAACAGCCTG ATGAAAGGTG 240
GTCTCGCCCT TTGAGGCCGA TTCTTGGGCG CTTGGCGGCG TCGAAGCGAT GCTCCACTAC 300
CGATTAAGAT AATTAAAATA AGGAAACCGC ATGGTTTCTT ATGTGAATTT GTCTGGCATA 360
CTCCAGCTCA AGGGCAATTT TTGGGCTATT GGCTGAGCAG TTGCCTCTAT ATGGTTATTC 420
AGAATAACAA TTGACTCCTC AGGAGGTCAG CGATGAGCAT TCTTGGTTTG AATGGTGCCC 480
CGGTCGGAGC TGAGCAGCTG GGCTCGGCTC TTGATCGCAT GAAGAAGGCG CACCTGGAGC 540
AGGGGCCTGC AAACTTGGAG CTGCGTCTGA GTAGGCTGGA TCGTGCGATT GCAATGCTTC 600
TGGAAAATCG TGAAGCAATT GCCGACGCGG TTTCTGCTGA CTTTGGCAAT CGCAGCCGTG 660
AGCAAACACT GCTTTGCGAC ATTGCTGGCT CGGTGGCAAG CCTGAAGGAT AGCCGCGAGC 720
ACGTGGCCAA ATGGATGGAG CCCGAACATC ACAAGGCGAT GTTTCCAGGG GCGGAGGCAC 780
GCGTTGAGTT TCAGCCGCTG GGTGTCGTTG GGGTCATTAG TCCCTGGAAC TTCCCTATCG 840
TACTGGCCTT TGGGCCGCTG GCCGGCATAT TCGCAGCAGG TAATCGCGCC ATGCTCAAGC 900
CGTCCGAGCT TACCCCGCGG ACTTCTGCCC TGCTTGCGGA GCTAATTGCT CGTTACTTCG 960
ATGAAACTGA GCTGACTACA GTGCTGGGCG ACGCTGAAGT CGGTGCGCTG TTCAGTGCTC 1020
AGCCTTTCGA TCATCTGATC TTCACCGGCG GCACTGCCGT GGCCAAGCAC ATCATGCGTG 1080
CCGCGGCGGA TAACCTAGTG CCCGTTACCC TGGAATTGGG TGGCAAATCG CCGGTGATCG 1140
TTTCCCGCAG TGCAGATATG GCGGACGTTG CACAACGGGT GTTGACGGTG AAAACCTTCA 1200
ATGCCGGGCA AATCTGTCTG GCACCGGACT ATGTGCTGGG GGAGAGGCGG TTTGCGTATT 1260
GGGCGCATGC ATAAAAACTG TTGTAATTCA TTAAGCATTC TGCCGACATG GAAGCCATCA 1320
CAAACGGCAT GATGAACCTG AATCGCCAGC GGCATCAGCA CCTTGTCGCC TTGCGTATAA 1380
TATTTGCCCA TGGACGCACA CCGTGGAAAC GGATGAAGGC ACGAACCCAG TTGACATAAG 1440
CCTGTTCGGT TCGTAAACTG TAATGCAAGT AGCGTATGCG CTCACGCAAC TGGTCCAGAA 1500
CCTTGACCGA ACGCAGCGGT GGTAACGGCG CAGTGGCGGT TTTCATGGCT TGTTATGACT 1560
GTTTTTTTGT ACAGTCTATG CCTCGGGCAT CCAAGCAGCA AGCGCGTTAC GCCGTGGGTC 1620
GATGTTTGAT GTTATGGAGC AGCAACGATG TTACGCAGCA GCAACGATGT TACGCAGCAG 1680
GGCAGTCGCC CTAAAACAAA GTTAGGTGGC TCAAGTATGG GCATCATTCG CACATGTAGG 1740
CTCGGCCCTG ACCAAGTCAA ATCCATGCGG GCTGCTCTTG ATCTTTTCGG TCGTGAGTTC 1800
GGAGACGTAG CCACCTACTC CCAACATCAG CCGGACTCCG ATTACCTCGG GAACTTGCTC 1860
CGTAGTAAGA CATTCATCGC GCTTGCTGCC TTCGACCAAG AAGCGGTTGT TGGCGCTCTC 1920
GCGGCTTACG TTCTGCCCAG GTTTGAGCAG CCGCGTAGTG AGATCTATAT CTATGATCTC 1980
GCAGTCTCCG GCGAGCACCG GAGGCAGGGC ATTGCCACCG CGCTCATCAA TCTCCTCAAG 2040
CATGAGGCCA ACGCGCTTGG TGCTTATGTG ATCTACGTGC AAGCAGATTA CGGTGACGAT 2100
CCCGCAGTGG CTCTCTATAC AAAGTTGGGC ATACGGGAAG AAGTGATGCA CTTTGATATC 2160

GACCCAAGTA CCGCCACCTA ACAATTCGTT CAAGCCGAGA TCGGCTTCCC TGCAAAGTCC 2220
TGTGGGTGAG TCGAACTTGG CGATGCGCGC ACCCTACGGA GAAGCGATCC ACGGACTGCT 2280
CTCTGTCCTC CTTTCAACGG AGTGTTAGAA CCGTTGGTAG TGGTTTTGGA CGGGCCCAGG 2340
AGCATGCGCT TCTGGGCCCG TTTCTTGAGT ATTCATTGGA TAGTCACGCG TGGTAGCTTC 2400
GAGCCTGCAC AGCTGATGAG CACCCTGGAA GGCGCGCTGT ACGCGGACGA CTGGGTTCAT 2460
CTTCGCCATT CATGACGGAA CTCCGTTCCC CAGTACCGCG ATGACTATTT TGCCTCTTCC 2520
GATGTCCGAT TCCACGCCGC CTGACGCTAA GCGGGGGCGG GGGCGCCCGC ATCCCAGCCC 2580
AGACAGCAAC AAATGAGTAG GCTCTTGGAT GCCGCGGCGG CTGAGATTGG TAACGGCAAT 2640
TTCGTCAATG TGACGATGGA TTCGATTGCC CGTGCTGCCG GCGTCTCAAA AAAAACGCTG 2700
TACGTCTTGG TGGCGAGCAA GGAAGAACTC ATTTCCCGGT TAGTGGCTCG AGACATGTCC 2760
AACCTTGAGG AATTC 2775

Sequence 6

```
GAATTCCGCG TATCGCCCGG TTCTATCAGC GGGCCGCTTT CGAAAGTCAT GGTGTTAGCC   60
GGTAGGGTCT TTTTCTTGGC CATGCTTGTT GCCTGAACCT TCGTTGACAT AGGGCAGAGG  120
TGCGTTTGCC GCTTCGCTTC GCGATGAACC GCATCGAGAT GCTGAGGTCA GGATTTTTCC  180
TTAACTCGCG TAAGCATTCT GTCATTTTTT TGGTGGCTTT GAACAGCCTG ATGAAAGGTG  240
GTCTCGCCCT TTGAGGCCGA TTCTTGGGCG CTTGGCGGCG TCGAAGCGAT GCTCCACTAC  300
CGATTAAGAT AATTAAAATA AGGAAACCGC ATGGTTTCTT ATGTGAATTT GTCTGGCATA  360
CTCCAGCTCA AGGGCAATTT TTGGGCTATT GGCTGAGCAG TTGCCTCTAT ATGGTTATTC  420
AGAATAACAA TTGACTCCTC AGGAGGTCAG CGATGAGCAT TCTTGGTTTG AATGGTGCCC  480
CGGTCGGAGC TGAGCAGCTG GGCTCGGCTC TTGATCGCAT GAAGAAGGCG CACCTGGAGC  540
AGGGGCCTGC AAACTTGGAG CTGCGTCTGA GTAGGCTGGA TCGTGCGATT GCAATGCTTC  600
TGGAAAATCG TGAAGCAATT GCCGACGCGG TTTCTGCTGA CTTTGGCAAT CGCAGCCGTG  660
AGCAAACACT GCTTTGCGAC ATTGCTGGCT CGGTGGCAAG CCTGAAGGAT AGCCGCGAGC  720
ACGTGGCCAA ATGGATGGAG CCCGAACATC ACAAGGCGAT GTTTCCAGGG GCGGAGGCAC  780
GCGTTGAGTT TCAGCCGCTG GGTGTCGTTG GGGTCATTAG TCCCTGGAAC TTCCCTATCG  840
TACTGGCCTT TGGGCCGCTG GCCGGCATAT TCGCAGCAGG TAATCGCGCC ATGCTCAAGC  900
CGTCCGAGCT TACCCCGCGG ACTTCTGCCC TGCTTGCGGA GCTAATTGCT CGTTACTTCG  960
ATGAAACTGA GCTGACTACA GTGCTGGGCG ACGCTGAAGT CGGTGCGCTG TTCAGTGCTC 1020
AGCCTTTCGA TCATCTGATC TTCACCGGCG GCACTGCCGT GGCCAAGCAC ATCATGCGTG 1080
CCGCGGCGGA TAACCTAGTG CCCGTTACCC TGGAATTGGG TGGCAAATCG CCGGTGATCG 1140
TTTCCCGCAG TGCAGATATG GCGGACGTTG CACAACGGGT GTTGACGGTG AAAACCTTCA 1200
ATGCCGGGCA AATCTGTCTG GCACCGTGGG TGAGTCGAAC TTGGCGATGC GCGCACCCTA 1260
CGGAGAAGCG ATCCACGGAC TGCTCTCTGT CCTCCTTTCA ACGGAGTGTT AGAACCGTTG 1320
GTAGTGGTTT TGGACGGGCC CAGGAGCATG CGCTTCTGGG CCCGTTTCTT GAGTATTCAT 1380
TGGATAGTCA CGCGTGGTAG CTTCGAGCCT GCACAGCTGA TGAGCACCCT GGAAGGCGCG 1440
CTGTACGCGG ACGACTGGGT TCATCTTCGC CATTCATGAC GGAACTCCGT TCCCCAGTAC 1500
CGCGATGACT ATTTTGCCTC TTCCGATGTC CGATTCCACG CCGCCTGACG CTAAGCGGGG 1560
GCGGGGGCGC CCGCATCCCA GCCCAGACAG CAACAAATGA GTAGGCTCTT GGATGCCGCG 1620
GCGGCTGAGA TTGGTAACGG CAATTTCGTC AATGTGACGA TGGATTCGAT TGCCCGTGCT 1680
GCCGGCGTCT CAAAAAAAAC GCTGTACGTC TTGGTGGCGA GCAAGGAAGA ACTCATTTCC 1740
CGGTTAGTGG CTCGAGACAT GTCCAACCTT GAGGAATTC                        1779
```

Sequence 7

```
CTGCAGCCGA GCATCGATTG AGCACTTTAC CCAGCTGCGC TGGCTGACCA TTCAGAATGG    60
CCCGCGGCAC TATCCAATCT AAATCGATCT TCGGGCGCCG CGGGCATCAT GCCCGCGGCG   120
CTCGCCTCAT TTCAATCTCT AACTTGATAA AAACAGAGCT GTTCTCCGGT CTTGGTGGAT   180
CAAGGCCAGT CGCGGAGAGT CTCGAAGAGG AGAGTACAGT GAACGCCGAG TCCACATTGC   240
AACCGCAGGC ATCATCATGC TCTGCTCAGC CACGCTACCG CAGTGTGTCG ATTGGTCATC   300
CTCCGGTTGA GGTTACGCAA GACGCTGGAG GTATTGTCCG GATGCGTTCT CTCGAGGCGC   360
TTCTTCCCTT CCCGGGTCGA ATTCTTGAGC GTCTCGAGCA TTGGGCTAAG ACCCGTCCAG   420
AACAAACCTG CGTTGCTGCC AGGGCGGCAA ATGGGGAATG GCGTCGTATC AGCTACGCGG   480
AAATGTTCCA CAACGTCCGC GCCATCGCAC AGAGCTTGCT TCCTTACGGA CTATCGGCAG   540
AGCGTCCGCT GCTTATCGTC TCTGGAAATG ACCTGGAACA TCTTCAGCTG GCATTTGGGG   600
CTATGTATGC GGGCATTCCC TATTGCCCGG TGTCTCCTGC TTATTCACTG CTGTCGCAAG   660
ATTTGGCGAA GCTGCGTCAC ATCGTAGGTC TTCTGCAACC GGGACTGGTC TTTGCTGCCG   720
ATGCAGCACC TTTCCAGGGG ACAGCAAGCG AACCGGAATT GCCAGCTGGG GCGCCCTCTG   780
GTAAGGTTGG GAAGCCCTGC AAAGTAAACT GGATGGCTTT CTTGCCGCCA AGGATCTGAT   840
GGCGCAGGGG ATCAAGATCT GATCAAGAGA CAGGATGAGG ATCGTTTCGC ATGATTGAAC   900
AAGATGGATT GCACGCAGGT TCTCCGGCCG CTTGGGTGGA GAGGCTATTC GGCTATGACT   960
GGGCACAACA GACAATCGGC TGCTCTGATG CCGCCGTGTT CCGGCTGTCA GCGCAGGGGC  1020
GCCCGGTTCT TTTTGTCAAG ACCGACCTGT CCGGTGCCCT GAATGAACTG CAGGACGAGG  1080
CAGCGCGGCT ATCGTGGCTG GCCACGACGG GCGTTCCTTG CGCAGCTGTG CTCGACGTTG  1140
TCACTGAAGC GGGAAGGGAC TGGCTGCTAT TGGGCGAAGT GCCGGGGCAG GATCTCCTGT  1200
CATCTCACCT TGCTCCTGCC GAGAAAGTAT CCATCATGGC TGATGCAATG CGGCGGCTGC  1260
ATACGCTTGA TCCGGCTACC TGCCCATTCG ACCACCAAGC GAAACATCGC ATCGAGCGAG  1320
CACGTACTCG GATGGAAGCC GGTCTTGTCG ATCAGGATGA TCTGGACGAA GAGCATCAGG  1380
GGCTCGCGCC AGCCGAACTG TTCGCCAGGC TCAAGGCGCG CATGCCCGAC GGCGAGGATC  1440
TCGTCGTGAC CCATGGCGAT GCCTGCTTGC CGAATATCAT GGTGGAAAAT GGCCGCTTTT  1500
CTGGATTCAT CGACTGTGGC CGGCTGGGTG TGGCGGACCG CTATCAGGAC ATAGCGTTGG  1560
CTACCCGTGA TATTGCTGAA GAGCTTGGCG GCGAATGGGC TGACCGCTTC CTCGTGCTTT  1620
ACGGTATCGC CGCTCCCGAT TCGCAGCGCA TCGCCTTCTA TCGCCTTCTT GACGAGTTCT  1680
TCTGAGCGGG ACTCTGGGGT TCGAAATGAC CGACCAAGCG ACGCCCCTGT TTTGCAATGG  1740
CGGTCGGCGA AAGTTGATGC GCTGTATCGT GGTGAAGATC AATCCATGCT GCGTGACGAG  1800
GCCACACTGT GAGTTGGTCA GGGGGGGCTT ACTCGGCGTT TTCCGACACT GCGTTGGTTG  1860
CGGCAGTGCG CACCCCCTGG ATTGATTGCG GGGGTGCCCT GTCGCTGGTG TCGCCTATCG  1920
ACTTAGGGGT AAAGGTCGCT CGCGAAGTTC TGATGCGTGC GTCGCTTGAA CCACAAATGG  1980
TCGATAGCGT ACTCGCAGGC TCTATGGCTC AAGCAAGCTT TGATGCTTAC CTGCTCCCGC  2040
GGCACATTGG CTTGTACAGC GGTGTTCCCA AGTCGGTTCC GGCCTTGGGG GTGCAGCGCA  2100
TTTGCGGCAC AGGCTTCGAA CTGCTTCGGC AGGCCGGCGA GCAGATTTCC CAAGGCGCTG  2160
ATCACGTGCT GTGTGTCGCG GGCTGCAG                                      2188
```

Sequence 8

```
CTGCAGCCGA GCATCGATTG AGCACTTTAC CCAGCTGCGC TGGCTGACCA TTCAGAATGG     60
CCCGCGGCAC TATCCAATCT AAATCGATCT TCGGGCGCCG CGGGCATCAT GCCCGCGGCG    120
CTCGCCTCAT TTCAATCTCT AACTTGATAA AAACAGAGCT GTTCTCCGGT CTTGGTGGAT    180
CAAGGCCAGT CGCGGAGAGT CTCGAAGAGG AGAGTACAGT GAACGCCGAG TCCACATTGC    240
AACCGCAGGC ATCATCATGC TCTGCTCAGC CACGCTACCG CAGTGTGTCG ATTGGTCATC    300
CTCCGGTTGA GGTTACGCAA GACGCTGGAG GTATTGTCCG GATGCGTTCT CTCGAGGCGC    360
TTCTTCCCTT CCCGGGTCGA ATTCTTGAGC GTCTCGAGCA TTGGGCTAAG ACCCGTCCAG    420
AACAAACCTG CGTTGCTGCC AGGGCGGCAA ATGGGGAATG GCGTCGTATC AGCTACGCGG    480
AAATGTTCCA CAACGTCCGC GCCATCGCAC AGAGCTTGCT TCCTTACGGA CTATCGGCAG    540
AGCGTCCGCT GCTTATCGTC TCTGGAAATG ACCTGGAACA TCTTCAGCTG GCATTTGGGG    600
CTATGTATGC GGGCATTCCC TATTGCCCGG TGTCTCCTGC TTATTCACTG CTGTCGCAAG    660
ATTTGGCGAA GCTGCGTCAC ATCGTAGGTC TTCTGCAACC GGGACTGGTC TTTGCTGCCG    720
ATGCAGCACC TTTCCAGGGG GAGAGGCGGT TTGCGTATTG GGCGCATGCA TAAAAACTGT    780
TGTAATTCAT TAAGCATTCT GCCGACATGG AAGCCATCAC AAACGGCATG ATGAACCTGA    840
ATCGCCAGCG GCATCAGCAC CTTGTCGCCT TGCGTATAAT ATTTGCCCAT GGACGCACAC    900
CGTGGAAACG GATGAAGGCA CGAACCCAGT TGACATAAGC CTGTTCGGTT CGTAAACTGT    960
AATGCAAGTA GCGTATGCGC TCACGCAACT GGTCCAGAAC CTTGACCGAA CGCAGCGGTG   1020
GTAACGGCGC AGTGGCGGTT TTCATGGCTT GTTATGACTG TTTTTTTGTA CAGTCTATGC   1080
CTCGGGCATC CAAGCAGCAA GCGCGTTACG CCGTGGGTCG ATGTTTGATG TTATGGAGCA   1140
GCAACGATGT TACGCAGCAG CAACGATGTT ACGCAGCAGG GCAGTCGCCC TAAAACAAAG   1200
TTAGGTGGCT CAAGTATGGG CATCATTCGC ACATGTAGGC TCGGCCCTGA CCAAGTCAAA   1260
TCCATGCGGG CTGCTCTTGA TCTTTTCGGT CGTGAGTTCG GAGACGTAGC CACCTACTCC   1320
CAACATCAGC CGGACTCCGA TTACCTCGGG AACTTGCTCC GTAGTAAGAC ATTCATCGCG   1380
CTTGCTGCCT TCGACCAAGA AGCGGTTGTT GGCGCTCTCG CGGCTTACGT TCTGCCCAGG   1440
TTTGAGCAGC CGCGTAGTGA GATCTATATC TATGATCTCG CAGTCTCCGG CGAGCACCGG   1500
AGGCAGGGCA TTGCCACCGC GCTCATCAAT CTCCTCAAGC ATGAGGCCAA CGCGCTTGGT   1560
GCTTATGTGA TCTACGTGCA AGCAGATTAC GGTGACGATC CCGCAGTGGC TCTCTATACA   1620
AAGTTGGGCA TACGGGAAGA AGTGATGCAC TTTGATATCG ACCCAAGTAC CGCCACCTAA   1680
CAATTCGTTC AAGCCGAGAT CGGCTTCCCC TGTTTTGCAA TGGCGGTCGG CGAAAGTTGA   1740
TGCGCTGTAT CGTGGTGAAG ATCAATCCAT GCTGCGTGAC GAGGCCACAC TGTGAGTTGG   1800
TCAGGGGGGG CTTACTCGGC GTTTTCCGAC ACTGCGTTGG TTGCGGCAGT GCGCACCCCC   1860
TGGATTGATT GCGGGGGTGC CCTGTCGCTG GTGTCGCCTA TCGACTTAGG GGTAAAGGTC   1920
GCTCGCGAAG TTCTGATGCG TGCGTCGCTT GAACCACAAA TGGTCGATAG CGTACTCGCA   1980
GGCTCTATGG CTCAAGCAAG CTTTGATGCT TACCTGCTCC CGCGGCACAT TGGCTTGTAC   2040
AGCGGTGTTC CCAAGTCGGT TCCGGCCTTG GGGGTGCAGC GCATTTGCGG CACAGGCTTC   2100
GAACTGCTTC GGCAGGCCGG CGAGCAGATT TCCCAAGGCG CTGATCACGT GCTGTGTGTC   2160
GCGGGCTGCA G                                                        2171
```

Sequence 9

CTGCAGCCGA	GCATCGATTG	AGCACTTTAC	CCAGCTGCGC	TGGCTGACCA	TTCAGAATGG	60
CCCGCGGCAC	TATCCAATCT	AAATCGATCT	TCGGGCGCCG	CGGGCATCAT	GCCCGCGGCG	120
CTCGCCTCAT	TTCAATCTCT	AACTTGATAA	AAACAGAGCT	GTTCTCCGGT	CTTGGTGGAT	180
CAAGGCCAGT	CGCGGAGAGT	CTCGAAGAGG	AGAGTACAGT	GAACGCCGAG	TCCACATTGC	240
AACCGCAGGC	ATCATCATGC	TCTGCTCAGC	CACGCTACCG	CAGTGTGTCG	ATTGGTCATC	300
CTCCGGTTGA	GGTTACGCAA	GACGCTGGAG	GTATTGTCCG	GATGCGTTCT	CTCGAGGCGC	360
TTCTTCCCTT	CCCGGGTCGA	ATTCTTGAGC	GTCTCGAGCA	TTGGGCTAAG	ACCCGTCCAG	420
AACAAACCTG	CGTTGCTGCC	AGGGCGGCAA	ATGGGGAATG	GCGTCGTATC	AGCTACGCGG	480
AAATGTTCCA	CAACGTCCGC	GCCATCGCAC	AGAGCTTGCT	TCCTTACGGA	CTATCGGCAG	540
AGCGTCCGCT	GCTTATCGTC	TCTGGAAATG	ACCTGGAACA	TCTTCAGCTG	GCATTTGGGG	600
CTATGTATGC	GGGCATTCCC	TATTGCCCGG	TGTCTCCTGC	TTATTCACTG	CTGTCGCAAG	660
ATTTGGCGAA	GCTGCGTCAC	ATCGTAGGTC	TTCTGCAACC	GGGACTGGTC	TTTGCTGCCG	720
ATGCAGCACC	TTTCCAGCGC	GCTGTTTTGC	AATGGCGGTC	GGCGAAAGTT	GATGCGCTGT	780
ATCGTGGTGA	AGATCAATCC	ATGCTGCGTG	ACGAGGCCAC	ACTGTGAGTT	GGTCAGGGGG	840
GGCTTACTCG	GCGTTTTCCG	ACACTGCGTT	GGTTGCGGCA	GTGCGCACCC	CCTGGATTGA	900
TTGCGGGGGT	GCCCTGTCGC	TGGTGTCGCC	TATCGACTTA	GGGGTAAAGG	TCGCTCGCGA	960
AGTTCTGATG	CGTGCGTCGC	TTGAACCACA	AATGGTCGAT	AGCGTACTCG	CAGGCTCTAT	1020
GGCTCAAGCA	AGCTTTGATG	CTTACCTGCT	CCCGCGGCAC	ATTGGCTTGT	ACAGCGGTGT	1080
TCCCAAGTCG	GTTCCGGCCT	TGGGGGTGCA	GCGCATTTGC	GGCACAGGCT	TCGAACTGCT	1140
TCGGCAGGCC	GGCGAGCAGA	TTTCCCAAGG	CGCTGATCAC	GTGCTGTGTG	TCGCGGGCTG	1200
CAG						1203

Sequence 10

```
GAATTCCCCT GGCGACGAAA GGGCGGCAGG CCGCATGGCC ACGGCTGGGC GGTAACTGAT    60
GCTTGCGTTA ATCGTTAACC GTTTGAAATT CCTTGCCAAA TTTCGGCGAG AGAATCATGC   120
GGGTACGCCT TTCCGTGCGC TTTGATCTGC GCTTCCGTGC CTTGAATCAG AAAAATAGTT   180
AATTGACAGA ACTATAGGTT CGCAGTAGCT TTTGCTCACC CACCAAATCC ACAGCACTGG   240
GGTGCACGAT GAATAGCTAC GATGGCCGTT GGTCTACCGT TGATGTGAAG GTTGAAGAAG   300
GTATCGCTTG GGTCACGCTG AACCGCCCGG AGAAGCGCAA CGCAATGAGC CCAACTCTCA   360
ATCGAGAGAT GGTCGAGGTT CTGGAGGTGC TGGAGCAGGA CGCAGATGCT CGCGTGCTTG   420
TTCTGACTGG TGCAGGCGAA TCCTGGACCG CGGGCATGGA CCTGAAGGAG TATTTCCGCG   480
AGACCGATGC TGGCCCCGAA ATTCTGCAAG AGAAGATTCG TCGGGGACAG CAAGCGAACC   540
GGAATTGCCA GCTGGGGCGC CCTCTGGTAA GGTTGGGAAG CCCTGCAAAG TAAACTGGAT   600
GGCTTTCTTG CCGCCAAGGA TCTGATGGCG CAGGGGATCA AGATCTGATC AAGAGACAGG   660
ATGAGGATCG TTTCGCATGA TTGAACAAGA TGGATTGCAC GCAGGTTCTC CGGCCGCTTG   720
GGTGGAGAGG CTATTCGGCT ATGACTGGGC ACAACAGACA ATCGGCTGCT CTGATGCCGC   780
CGTGTTCCGG CTGTCAGCGC AGGGGCGCCC GGTTCTTTTT GTCAAGACCG ACCTGTCCGG   840
TGCCCTGAAT GAACTGCAGG ACGAGGCAGC GCGGCTATCG TGGCTGGCCA CGACGGGCGT   900
TCCTTGCGCA GCTGTGCTCG ACGTTGTCAC TGAAGCGGGA AGGGACTGGC TGCTATTGGG   960
CGAAGTGCCG GGGCAGGATC TCCTGTCATC TCACCTTGCT CCTGCCGAGA AAGTATCCAT  1020
CATGGCTGAT GCAATGCGGC GGCTGCATAC GCTTGATCCG GCTACCTGCC CATTCGACCA  1080
CCAAGCGAAA CATCGCATCG AGCGAGCACG TACTCGGATG GAAGCCGGTC TTGTCGATCA  1140
GGATGATCTG GACGAAGAGC ATCAGGGGCT CGCGCCAGCC GAACTGTTCG CCAGGCTCAA  1200
GGCGCGCATG CCCGACGGCG AGGATCTCGT CGTGACCCAT GGCGATGCCT GCTTGCCGAA  1260
TATCATGGTG GAAAATGGCC GCTTTTCTGG ATTCATCGAC TGTGGCCGGC TGGGTGTGGC  1320
GGACCGCTAT CAGGACATAG CGTTGGCTAC CCGTGATATT GCTGAAGAGC TTGGCGGCGA  1380
ATGGGCTGAC CGCTTCCTCG TGCTTTACGG TATCGCCGCT CCCGATTCGC AGCGCATCGC  1440
CTTCTATCGC CTTCTTGACG AGTTCTTCTG AGCGGGACTC TGGGGTTCGA AATGACCGAC  1500
CAAGCGACGC CCCGAGCAGG GCATGAAGCA GTTCCTTGAC GAGAAAAGCA TCAAGCCGGG  1560
CTTGCAGACC TACAAGCGCT GATAAATGCG CCGGGGCCCT CGCTGCGCCC CCGGCCTTCC  1620
AATAATGACA ATAATGAGGA GTGCCCAATG TTTCACGTGC CCCTGCTTAT TGGTGGTAAG  1680
CCTTGTTCAG CATCTGATGA GCGCACCTTC GAGCGTCGTA GCCCGCTGAC CGGAGAAGTG  1740
GTATCGCGCG TCGCTGCTGC CAGTTTGGAA GATGCGGACG CCGCAGTGGC CGCTGCACAG  1800
GCTGCGTTTC CTGAATGGGC GGCGCTTGCT CCGAGCGAAC GCCGTGCCCG ACTGCTGCGA  1860
GCGGCGGATC TTCTAGAGGA CCGTTCTTCC GAGTTCACCG CCGCAGCGAG TGAAACTGGC  1920
GCAGCGGGAA ACTGGTATGG GTTTAACGTT TACCTGGCGG CGGGCATGTT GCGGGGAATT  1980
C                                                                   1981
```

Sequence 11

```
GAATTCCCCT GGCGACGAAA GGGCGGCAGG CCGCATGGCC ACGGCTGGGC GGTAACTGAT    60
GCTTGCGTTA ATCGTTAACC GTTTGAAATT CCTTGCCAAA TTTCGGCGAG AGAATCATGC   120
GGGTACGCCT TTCCGTGCGC TTTGATCTGC GCTTCCGTGC CTTGAATCAG AAAAATAGTT   180
AATTGACAGA ACTATAGGTT CGCAGTAGCT TTTGCTCACC CACCAAATCC ACAGCACTGG   240
GGTGCACGAT GAATAGCTAC GATGGCCGTT GGTCTACCGT TGATGTGAAG GTTGAAGAAG   300
GTATCGCTTG GGTCACGCTG AACCGCCCGG AGAAGCGCAA CGCAATGAGC CCAACTCTCA   360
ATCGAGAGAT GGTCGAGGTT CTGGAGGTGC TGGAGCAGGA CGCAGATGCT CGCGTGCTTG   420
TTCTGACTGG TGCAGGCGAA TCCTGGACCG CGGGCATGGA CCTGAAGGAG TATTTCCGCG   480
AGACCGATGC TGGCCCCGAA ATTCTGCAAG AGAAGATTCG TCGGGGGAGA GGCGGTTTGC   540
GTATTGGGCG CATGCATAAA AACTGTTGTA ATTCATTAAG CATTCTGCCG ACATGGAAGC   600
CATCACAAAC GGCATGATGA ACCTGAATCG CCAGCGGCAT CAGCACCTTG TCGCCTTGCG   660
TATAATATTT GCCCATGGAC GCACACCGTG GAAACGGATG AAGGCACGAA CCCAGTTGAC   720
ATAAGCCTGT TCGGTTCGTA AACTGTAATG CAAGTAGCGT ATGCGCTCAC GCAACTGGTC   780
CAGAACCTTG ACCGAACGCA GCGGTGGTAA CGGCGCAGTG GCGGTTTTCA TGGCTTGTTA   840
TGACTGTTTT TTTGTACAGT CTATGCCTCG GGCATCCAAG CAGCAAGCGC GTTACGCCGT   900
GGGTCGATGT TTGATGTTAT GGAGCAGCAA CGATGTTACG CAGCAGCAAC GATGTTACGC   960
AGCAGGGCAG TCGCCCTAAA ACAAAGTTAG GTGGCTCAAG TATGGGCATC ATTCGCACAT  1020
GTAGGCTCGG CCCTGACCAA GTCAAATCCA TGCGGGCTGC TCTTGATCTT TTCGGTCGTG  1080
AGTTCGGAGA CGTAGCCACC TACTCCCAAC ATCAGCCGGA CTCCGATTAC CTCGGGAACT  1140
TGCTCCGTAG TAAGACATTC ATCGCGCTTG CTGCCTTCGA CCAAGAAGCG GTTGTTGGCG  1200
CTCTCGCGGC TTACGTTCTG CCCAGGTTTG AGCAGCCGCG TAGTGAGATC TATATCTATG  1260
ATCTCGCAGT CTCCGGCGAG CACCGGAGGC AGGGCATTGC CACCGCGCTC ATCAATCTCC  1320
TCAAGCATGA GGCCAACGCG CTTGGTGCTT ATGTGATCTA CGTGCAAGCA GATTACGGTG  1380
ACGATCCCGC AGTGGCTCTC TATACAAAGT TGGGCATACG GGAAGAAGTG ATGCACTTTG  1440
ATATCGACCC AAGTACCGCC ACCTAACAAT TCGTTCAAGC CGAGATCGGC TTCCCCGAGC  1500
AGGGCATGAA GCAGTTCCTT GACGAGAAAA GCATCAAGCC GGGCTTGCAG ACCTACAAGC  1560
GCTGATAAAT GCGCCGGGGC CCTCGCTGCG CCCCCGGCCT TCCAATAATG ACAATAATGA  1620
GGAGTGCCCA ATGTTTCACG TGCCCCTGCT TATTGGTGGT AAGCCTTGTT CAGCATCTGA  1680
TGAGCGCACC TTCGAGCGTC GTAGCCCGCT GACCGGAGAA GTGGTATCGC GCGTCGCTGC  1740
TGCCAGTTTG GAAGATGCGG ACGCCGCAGT GGCCGCTGCA CAGGCTGCGT TTCCTGAATG  1800
GGCGGCGCTT GCTCCGAGCG AACGCCGTGC CCGACTGCTG CGAGCGGCGG ATCTTCTAGA  1860
GGACCGTTCT TCCGAGTTCA CCGCCGCAGC GAGTGAAACT GGCGCAGCGG GAAACTGGTA  1920
TGGGTTTAAC GTTTACCTGG CGGCGGGCAT GTTGCGGGGA ATTC                    1964
```

Sequence 12

```
GAATTCCCCT GGCGACGAAA GGGCGGCAGG CCGCATGGCC ACGGCTGGGC GGTAACTGAT    60
GCTTGCGTTA ATCGTTAACC GTTTGAAATT CCTTGCCAAA TTTCGGCGAG AGAATCATGC   120
GGGTACGCCT TTCCGTGCGC TTTGATCTGC GCTTCCGTGC CTTGAATCAG AAAAATAGTT   180
AATTGACAGA ACTATAGGTT CGCAGTAGCT TTTGCTCACC CACCAAATCC ACAGCACTGG   240
GGTGCACGAT GAATAGCTAC GATGGCCGTT GGTCTACCGT TGATGTGAAG GTTGAAGAAG   300
GTATCGCTTG GGTCACGCTG AACCGCCCGG AGAAGCGCAA CGCAATGAGC CCAACTCTCA   360
ATCGAGAGAT GGTCGAGGTT CTGGAGGTGC TGGAGCAGGA CGCAGATGCT CGCGTGCTTG   420
TTCTGACTGG TGCAGGCGAA TCCTGGACCG CGGGCATGGA CCTGAAGGAG TATTTCCGCG   480
AGACCGATGC TGGCCCCGAA ATTCTGCAAG AGAAGATTCG TCGCGAGCAG GGCATGAAGC   540
AGTTCCTTGA CGAGAAAAGC ATCAAGCCGG GCTTGCAGAC CTACAAGCGC TGATAAATGC   600
GCCGGGGCCC TCGCTGCGCC CCCGGCCTTC CAATAATGAC AATAATGAGG AGTGCCCAAT   660
GTTTCACGTG CCCCTGCTTA TTGGTGGTAA GCCTTGTTCA GCATCTGATG AGCGCACCTT   720
CGAGCGTCGT AGCCCGCTGA CCGGAGAAGT GGTATCGCGC GTCGCTGCTG CCAGTTTGGA   780
AGATGCGGAC GCCGCAGTGG CCGCTGCACA GGCTGCGTTT CCTGAATGGG CGGCGCTTGC   840
TCCGAGCGAA CGCCGTGCCC GACTGCTGCG AGCGGCGGAT CTTCTAGAGG ACCGTTCTTC   900
CGAGTTCACC GCCGCAGCGA GTGAAACTGG CGCAGCGGGA AACTGGTATG GGTTTAACGT   960
TTACCTGGCG GCGGGCATGT TGCGGGGAAT TC                                 992
```

Sequence 13

```
GAATTCCAAT AATGACAATA ATGAGGAGTG CCCAATGTTT CACGTGCCCC TGCTTATTGG     60
TGGTAAGCCT TGTTCAGCAT CTGATGAGCG CACCTTCGAG CGTCGTAGCC CGCTGACCGG    120
AGAAGTGGTA TCGCGCGTCG CTGCTGCCAG TTTGGAAGAT GCGGACGCCG CAGTGGCCGC    180
TGCACAGGCT GCGTTTCCTG AATGGGCGGC GCTTGCTCCG AGCGAACGCC GTGCCCGACT    240
GCTGCGAGCG GCGGATCTTC TAGAGGACCG TTCTTCCGAG TTCACCGCCG CAGCGAGTGA    300
AACTGGCGCA GCGGGAAACT GGTATGGGTT TAACGTTTAC CTGGCGGCGG GCATGTTGCG    360
GGAAGCCGCG GCCATGACCA CACAGATTCA GGGCGATGTC ATTCCGTCCA ATGTGCCCGG    420
TAGCTTTGCC ATGGCGGTTC GACAGCCATG TGGCGTGGTG CTCGGTATTG CGCCTTGGAA    480
TGCTCCGGTA ATCCTTGGCG TACGGGCTGT TGCGATGCCG TTGGCATGCG GCAATACCGT    540
GGTGTTGAAA AGCTCTGAGC TGAGTCCCTT TACCCATCGC CTGATTGGTC AGGTGTTGCA    600
TGATGCTGGT CTGGGGGATG GCGTGGTGAA TGTCATCAGC AATGCCCCGC AAGACGCTCC    660
TGCGGTGGTG GAGCGACTGA TTGCAAATCC TGCGGTACGT CGAGTGAACT TCACCGGTTC    720
GACCCACGTT GGACGGATCA TTGGTGAGCT GTCTGCGCGT CATCTGAAGC CTGCTGTGCT    780
GGAATTAGGT GGTAAGGCTC CGTTCTTGGT CTTGGACGAT GCCGACCTCG ATGCGGCGGT    840
CGAAGCGGCG GCCTTTGGTG CCTACTTCAA TCAGGGTCAA ATCTGCATGT CCACTGAGCG    900
TCTGATTGTG ACAGCAGTCG CAGACGCCTT TGTTGAAAAG CTGGCGAGGA AGGTCGCCAC    960
ACTGCGTGCT GGCGATCCTA ATGATCCGCA ATCGGTCTTG GGTTCGTTGA TTGATGCCAA   1020
TGCAGGTCAA CGCATCCAGG TTCTGGTCGA TGATGCGCTC GGGGACAGCA AGCGAACCGG   1080
AATTGCCAGC TGGGGCGCCC TCTGGTAAGG TTGGGAAGCC CTGCAAAGTA AACTGGATGG   1140
CTTTCTTGCC GCCAAGGATC TGATGGCGCA GGGGATCAAG ATCTGATCAA GAGACAGGAT   1200
GAGGATCGTT TCGCATGATT GAACAAGATG GATTGCACGC AGGTTCTCCG GCCGCTTGGG   1260
TGGAGAGGCT ATTCGGCTAT GACTGGGCAC AACAGACAAT CGGCTGCTCT GATGCCGCCG   1320
TGTTCCGGCT GTCAGCGCAG GGGCGCCCGG TTCTTTTTGT CAAGACCGAC CTGTCCGGTG   1380
CCCTGAATGA ACTGCAGGAC GAGGCAGCGC GGCTATCGTG GCTGGCCACG ACGGGCGTTC   1440
CTTGCGCAGC TGTGCTCGAC GTTGTCACTG AAGCGGGAAG GGACTGGCTG CTATTGGGCG   1500
AAGTGCCGGG GCAGGATCTC CTGTCATCTC ACCTTGCTCC TGCCGAGAAA GTATCCATCA   1560
TGGCTGATGC AATGCGGCGG CTGCATACGC TTGATCCGGC TACCTGCCCA TTCGACCACC   1620
AAGCGAAACA TCGCATCGAG CGAGCACGTA CTCGGATGGA AGCCGGTCTT GTCGATCAGG   1680
ATGATCTGGA CGAAGAGCAT CAGGGGCTCG CGCCAGCCGA ACTGTTCGCC AGGCTCAAGG   1740
CGCGCATGCC CGACGGCGAG GATCTCGTCG TGACCCATGG CGATGCCTGC TTGCCGAATA   1800
TCATGGTGGA AAATGGCCGC TTTTCTGGAT TCATCGACTG TGGCCGGCTG GGTGTGGCGG   1860
ACCGCTATCA GGACATAGCG TTGGCTACCC GTGATATTGC TGAAGAGCTT GGCGGCGAAT   1920
GGGCTGACCG CTTCCTCGTG CTTTACGGTA TCGCCGCTCC CGATTCGCAG CGCATCGCCT   1980
TCTATCGCCT TCTTGACGAG TTCTTCTGAG CGGGACTCTG GGGTTCGAAA TGACCGACCA   2040
AGCGACGCCC GGCCCAGCGC GTCGATTCGG GCATTTGCCA TATCAATGGA CCGACTGTGC   2100
ATGACGAGGC TCAGATGCCA TTCGGTGGGG TGAAGTCCAG CGGCTACGGC AGCTTCGGCA   2160
```

GTCGAGCATC GATTGAGCAC TTTACCCAGC TGCGCTGGCT GACCATTCAG AATGGCCCGC 2220
GGCACTATCC AATCTAAATC GATCTTCGGG CGCCGCGGGC ATCATGCCCG CGGCGCTCGC 2280
CTCATTTCAA TCTCTAACTT GATAAAAACA GAGCTGTTCT CCGGTCTTGG TGGATCAAGG 2340
CCAGTCGCGG AGAGTCTCGA AGAGGAGAGT ACAGTGAACG CCGAGTCCAC ATTGCAACCG 2400
CAGGCATCAT CATGCTCTGC TCAGCCACGC TACCGCAGTG TGTCGATTGG TCATCCTCCG 2460
GTTGAGGTTA CGCAAGACGC TGGAGGTATT GTCCGGATGC GTTCTCTCGA GGCGCTTCTT 2520
CCCTTCCCGG GTGGAATTC 2539

Sequence 14

```
GAATTCCAAT AATGACAATA ATGAGGAGTG CCCAATGTTT CACGTGCCCC TGCTTATTGG   60
TGGTAAGCCT TGTTCAGCAT CTGATGAGCG CACCTTCGAG CGTCGTAGCC CGCTGACCGG  120
AGAAGTGGTA TCGCGCGTCG CTGCTGCCAG TTTGGAAGAT GCGGACGCCG CAGTGGCCGC  180
TGCACAGGCT GCGTTTCCTG AATGGGCGGC GCTTGCTCCG AGCGAACGCC GTGCCCGACT  240
GCTGCGAGCG GCGGATCTTC TAGAGGACCG TTCTTCCGAG TTCACCGCCG CAGCGAGTGA  300
AACTGGCGCA GCGGGAAACT GGTATGGGTT TAACGTTTAC CTGGCGGCGG GCATGTTGCG  360
GGAAGCCGCG GCCATGACCA CACAGATTCA GGGCGATGTC ATTCCGTCCA ATGTGCCCGG  420
TAGCTTTGCC ATGGCGGTTC GACAGCCATG TGGCGTGGTG CTCGGTATTG CGCCTTGGAA  480
TGCTCCGGTA ATCCTTGGCG TACGGGCTGT TGCGATGCCG TTGGCATGCG GCAATACCGT  540
GGTGTTGAAA AGCTCTGAGC TGAGTCCCTT TACCCATCGC CTGATTGGTC AGGTGTTGCA  600
TGATGCTGGT CTGGGGGATG GCGTGGTGAA TGTCATCAGC AATGCCCCGC AAGACGCTCC  660
TGCGGTGGTG GAGCGACTGA TTGCAAATCC TGCGGTACGT CGAGTGAACT TCACCGGTTC  720
GACCCACGTT GGACGGATCA TTGGTGAGCT GTCTGCGCGT CATCTGAAGC CTGCTGTGCT  780
GGAATTAGGT GGTAAGGCTC CGTTCTTGGT CTTGGACGAT GCCGACCTCG ATGCGGCGGT  840
CGAAGCGGCG GCCTTTGGTG CCTACTTCAA TCAGGGTCAA ATCTGCATGT CCACTGAGCG  900
TCTGATTGTG ACAGCAGTCG CAGACGCCTT TGTTGAAAAG CTGGCGAGGA AGGTCGCCAC  960
ACTGCGTGCT GGCGATCCTA ATGATCCGCA ATCGGTCTTG GGTTCGTTGA TTGATGCCAA 1020
TGCAGGTCAA CGCATCCAGG TGGGGAGAGG CGGTTTGCGT ATTGGGCGCA TGCATAAAAA 1080
CTGTTGTAAT TCATTAAGCA TTCTGCCGAC ATGGAAGCCA TCACAAACGG CATGATGAAC 1140
CTGAATCGCC AGCGGCATCA GCACCTTGTC GCCTTGCGTA TAATATTTGC CCATGGACGC 1200
ACACCGTGGA AACGGATGAA GGCACGAACC CAGTTGACAT AAGCCTGTTC GGTTCGTAAA 1260
CTGTAATGCA AGTAGCGTAT GCGCTCACGC AACTGGTCCA GAACCTTGAC CGAACGCAGC 1320
GGTGGTAACG GCGCAGTGGC GGTTTTCATG GCTTGTTATG ACTGTTTTTT TGTACAGTCT 1380
ATGCCTCGGG CATCCAAGCA GCAAGCGCGT TACGCCGTGG GTCGATGTTT GATGTTATGG 1440
AGCAGCAACG ATGTTACGCA GCAGCAACGA TGTTACGCAG CAGGGCAGTC GCCCTAAAAC 1500
AAAGTTAGGT GGCTCAAGTA TGGGCATCAT TCGCACATGT AGGCTCGGCC CTGACCAAGT 1560
CAAATCCATG CGGGCTGCTC TTGATCTTTT CGGTCGTGAG TTCGGAGACG TAGCCACCTA 1620
CTCCCAACAT CAGCCGGACT CCGATTACCT CGGGAACTTG CTCCGTAGTA AGACATTCAT 1680
CGCGCTTGCT GCCTTCGACC AAGAAGCGGT TGTTGGCGCT CTCGCGGCTT ACGTTCTGCC 1740
CAGGTTTGAG CAGCCGCGTA GTGAGATCTA TATCTATGAT CTCGCAGTCT CCGGCGAGCA 1800
CCGGAGGCAG GGCATTGCCA CCGCGCTCAT CAATCTCCTC AAGCATGAGG CCAACGCGCT 1860
TGGTGCTTAT GTGATCTACG TGCAAGCAGA TTACGGTGAC GATCCCGCAG TGGCTCTCTA 1920
TACAAAGTTG GGCATACGGG AAGAAGTGAT GCACTTTGAT ATCGACCCAA GTACCGCCAC 1980
CTAACAATTC GTTCAAGCCG AGATCGGCTT CCCAATTGGC CCAGCGCGTC GATTCGGGCA 2040
TTTGCCATAT CAATGGACCG ACTGTGCATG ACGAGGCTCA GATGCCATTC GGTGGGGTGA 2100
AGTCCAGCGG CTACGGCAGC TTCGGCAGTC GAGCATCGAT TGAGCACTTT ACCCAGCTGC 2160
```

```
GCTGGCTGAC CATTCAGAAT GGCCCGCGGC ACTATCCAAT CTAAATCGAT CTTCGGGCGC  2220
CGCGGGCATC ATGCCCGCGG CGCTCGCCTC ATTTCAATCT CTAACTTGAT AAAAACAGAG  2280
CTGTTCTCCG GTCTTGGTGG ATCAAGGCCA GTCGCGGAGA GTCTCGAAGA GGAGAGTACA  2340
GTGAACGCCG AGTCCACATT GCAACCGCAG GCATCATCAT GCTCTGCTCA GCCACGCTAC  2400
CGCAGTGTGT CGATTGGTCA TCCTCCGGTT GAGGTTACGC AAGACGCTGG AGGTATTGTC  2460
CGGATGCGTT CTCTCGAGGC GCTTCTTCCC TTCCCGGGTG GAATTC                 2506
```

Sequence 15

```
GAATTCCAAT AATGACAATA ATGAGGAGTG CCCAATGTTT CACGTGCCCC TGCTTATTGG     60
TGGTAAGCCT TGTTCAGCAT CTGATGAGCG CACCTTCGAG CGTCGTAGCC CGCTGACCGG    120
AGAAGTGGTA TCGCGCGTCG CTGCTGCCAG TTTGGAAGAT GCGGACGCCG CAGTGGCCGC    180
TGCACAGGCT GCGTTTCCTG AATGGGCGGC GCTTGCTCCG AGCGAACGCC GTGCCCGACT    240
GCTGCGAGCG GCGGATCTTC TAGAGGACCG TTCTTCCGAG TTCACCGCCG CAGCGAGTGA    300
AACTGGCGCA GCGGGAAACT GGTATGGGTT TAACGTTTAC CTGGCGGCGG GCATGTTGCG    360
GGAAGCCGCG GCCATGACCA CACAGATTCA GGGCGATGTC ATTCCGTCCA ATGTGCCCGG    420
TAGCTTTGCC ATGGCGGTTC GACAGCCATG TGGCGTGGTG CTCGGTATTG CGCCTTGGAA    480
TGCTCCGGTA ATCCTTGGCG TACGGGCTGT TGCGATGCCG TTGGCATGCG GCAATACCGT    540
GGTGTTGAAA AGCTCTGAGC TGAGTCCCTT TACCCATCGC CTGATTGGTC AGGTGTTGCA    600
TGATGCTGGT CTGGGGGATG GCGTGGTGAA TGTCATCAGC AATGCCCCGC AAGACGCTCC    660
TGCGGTGGTG GAGCGACTGA TTGCAAATCC TGCGGTACGT CGAGTGAACT TCACCGGTTC    720
GACCCACGTT GGACGGATCA TTGGTGAGCT GTCTGCGCGT CATCTGAAGC CTGCTGTGCT    780
GGAATTAGGT GGTAAGGCTC CGTTCTTGGT CTTGGACGAT GCCGACCTCG ATGCGGCGGT    840
CGAAGCGGCG GCCTTTGGTG CCTACTTCAA TCAGGGTCAA ATCTGCATGT CCACTGAGCG    900
TCTGATTGTG ACAGCAGTCG CAGACGCCTT TGTTGAAAAG CTGGCGAGGA AGGTCGCCAC    960
ACTGCGTGCT GGCGATCCTA ATGATCCGCA ATCGGTCTTG GGTTCGTTGA TTGATGCCAA   1020
TGCAGGTCAA CGCATCCAGG TTCTGGTCGA TGATGCGCTC GCAAAAGGCG CGCAATGGAA   1080
TTGGCCCAGC GCGTCGATTC GGGCATTTGC CATATCAATG GACCGACTGT GCATGACGAG   1140
GCTCAGATGC CATTCGGTGG GGTGAAGTCC AGCGGCTACG GCAGCTTCGG CAGTCGAGCA   1200
TCGATTGAGC ACTTTACCCA GCTGCGCTGG CTGACCATTC AGAATGGCCC GCGGCACTAT   1260
CCAATCTAAA TCGATCTTCG GGCGCCGCGG GCATCATGCC CGCGGCGCTC GCCTCATTTC   1320
AATCTCTAAC TTGATAAAAA CAGAGCTGTT CTCCGGTCTT GGTGGATCAA GGCCAGTCGC   1380
GGAGAGTCTC GAAGAGGAGA GTACAGTGAA CGCCGAGTCC ACATTGCAAC CGCAGGCATC   1440
ATCATGCTCT GCTCAGCCAC GCTACCGCAG TGTGTCGATT GGTCATCCTC CGGTTGAGGT   1500
TACGCAAGAC GCTGGAGGTA TTGTCCGGAT GCGTTCTCTC GAGGCGCTTC TTCCCTTCCC   1560
GGGTGGAATT C                                                        1571
```

Sequence 16

```
GAATTCCGCG GTCGGCGAAA GTTGATGCGC TGTATCGTGG TGAAGATCAA TCCATGCTGC   60
GTGACGAGGC CACACTGTGA GTTGGTCAGG GGGGGCTTAC TCGGCGTTTT CCGACACTGC  120
GTTGGTTGCG GCAGTGCGCA CCCCCTGGAT TGATTGCGGG GGTGCCCTGT CGCTGGTGTC  180
GCCTATCGAC TTAGGGGTAA AGGTCGCTCG CGAAGTTCTG ATGCGTGCGT CGCTTGAACC  240
ACAAATGGTC GATAGCGTAC TCGCAGGCTC TATGGCTCAA GCAAGCTTTG ATGCTTACCT  300
GCTCCCGCGG CACATTGGCT TGTACAGCGG TGTTCCCAAG TCGGTTCCGG CCTTGGGGGT  360
GCAGCGCATT TGCGGCACAG GCTTCGAACT GCTTCGGCAG GCCGGCGAGC AGATTTCCCA  420
AGGCGCTGAT CACGTGCTGT GTGTCGCGGC AGAGTCCATG TCGCGTAACC CCATCGCGTC  480
GTATACACAC CGGGGCGGGT TCCGCCTCGG TGCGCCCGTT GAGTTCAAGG ATTTTTTGTG  540
GGAGGCATTG TTTGATCCTG CTCCAGGACT CGACATGATC GCTACCGCAG AAAACCTGGG  600
GACAGCAAGC GAACCGGAAT TGCCAGCTGG GGCGCCCTCT GGTAAGGTTG GGAAGCCCTG  660
CAAAGTAAAC TGGATGGCTT TCTTGCCGCC AAGGATCTGA TGGCGCAGGG GATCAAGATC  720
TGATCAAGAG ACAGGATGAG GATCGTTTCG CATGATTGAA CAAGATGGAT TGCACGCAGG  780
TTCTCCGGCC GCTTGGGTGG AGAGGCTATT CGGCTATGAC TGGGCACAAC AGACAATCGG  840
CTGCTCTGAT GCCGCCGTGT TCCGGCTGTC AGCGCAGGGG CGCCCGGTTC TTTTTGTCAA  900
GACCGACCTG TCCGGTGCCC TGAATGAACT GCAGGACGAG GCAGCGCGGC TATCGTGGCT  960
GGCCACGACG GGCGTTCCTT GCGCAGCTGT GCTCGACGTT GTCACTGAAG CGGGAAGGGA 1020
CTGGCTGCTA TTGGGCGAAG TGCCGGGGCA GGATCTCCTG TCATCTCACC TTGCTCCTGC 1080
CGAGAAAGTA TCCATCATGG CTGATGCAAT GCGGCGGCTG CATACGCTTG ATCCGGCTAC 1140
CTGCCCATTC GACCACCAAG CGAAACATCG CATCGAGCGA GCACGTACTC GGATGGAAGC 1200
CGGTCTTGTC GATCAGGATG ATCTGGACGA AGAGCATCAG GGGCTCGCGC CAGCCGAACT 1260
GTTCGCCAGG CTCAAGGCGC GCATGCCCGA CGGCGAGGAT CTCGTCGTGA CCCATGGCGA 1320
TGCCTGCTTG CCGAATATCA TGGTGGAAAA TGGCCGCTTT TCTGGATTCA TCGACTGTGG 1380
CCGGCTGGGT GTGGCGGACC GCTATCAGGA CATAGCGTTG GCTACCCGTG ATATTGCTGA 1440
AGAGCTTGGC GGCGAATGGG CTGACCGCTT CCTCGTGCTT TACGGTATCG CCGCTCCCGA 1500
TTCGCAGCGC ATCGCCTTCT ATCGCCTTCT TGACGAGTTC TTCTGAGCGG GACTCTGGGG 1560
TTCGAAATGA CCGACCAAGC GACGCCCATT GAGGGCGCAA GAGGAGAAAT GGATTGACCA 1620
AGAGATCGTG GCTGTTACGG ATGAACAGTT CGATTTAGAG GGCTACAACA GTCGAGCAAT 1680
TGAACTGCCT CGGAAGGCAA AATTGTTGAT CGTGACAGTC ATCCGCGGCC TAGCAGTCTT 1740
TGAAGCCCTT TCCCGATTGA AGCCTGTTCA TTCTGGCGGG GTGCAGACTG CGGGCAACAG 1800
CTGTGCCGTA GTGGACGGCG CCGCGGCGGC TTTGGTGGCT CGAGAGTCGT CTGCGACACA 1860
GCCGGTCTTG GCTAGGATAC TGGCTACCTC CGTAGTCGGG ATCGAGCCCG AGCATATGGG 1920
GCTCGGCCCT GCGCCCGCGA TTCGCCTGCT GCTTGCGCGT AGTGATCTTA GTTTGAGGGA 1980
TATCGACCTC TTTGAGATAA ACGAGGCGCA GGCCGCCCAA GTTCTAGCGG TACAGCATGA 2040
ATTGGGTATT GAGCACTCAA AACTTAATAT TTGGGGCGGG GCCATTGCAC TTGGACACCC 2100
GCTTGCCGCG ACCGGATTGC GTCTCTGCAT GACCCTCGCT CACCAATTGC AAGCTAATAA 2160
```

```
CTTTCGATAT GGAATTGCCT CGGCATGCAT TGGTGGGGGA CAGGGGATGG CGGTTCTTTT  2220
AGAGAATCCC CACTTCGGTT CGTCCTCTGC ACGAAGTTCG ATGATTAACA GAGTTGACCA  2280
CTATCCACTG AGCTAACGGG CATCTCCTTT GTTGCTTTGA GGTGGCGCAC GAAGGAGGGC  2340
TCGAAAATCT CTGCTAAAAA CAAGAAGAAG GAACAGGGAA CATGATTAGT TTCGCTCGTA  2400
TGGCAGAAAG TTTAGGAGTC CAGGCTAAAC TTGCCCTTGC CTTCGCACTC GTATTATGTG  2460
TCGGGCTGAT TGTTACCGGC ACGGGTTTCT ACAGTGTACA TACCTTGTCA GGGTTGGTGG  2520
GAATTC                                                             2526
```

Sequence 17

GAATTCCGCG GTCGGCGAAA GTTGATGCGC TGTATCGTGG TGAAGATCAA TCCATGCTGC 60
GTGACGAGGC CACACTGTGA GTTGGTCAGG GGGGGCTTAC TCGGCGTTTT CCGACACTGC 120
GTTGGTTGCG GCAGTGCGCA CCCCCTGGAT TGATTGCGGG GGTGCCCTGT CGCTGGTGTC 180
GCCTATCGAC TTAGGGGTAA AGGTCGCTCG CGAAGTTCTG ATGCGTGCGT CGCTTGAACC 240
ACAAATGGTC GATAGCGTAC TCGCAGGCTC TATGGCTCAA GCAAGCTTTG ATGCTTACCT 300
GCTCCCGCGG CACATTGGCT TGTACAGCGG TGTTCCCAAG TCGGTTCCGG CCTTGGGGGT 360
GCAGCGCATT TGCGGCACAG GCTTCGAACT GCTTCGGCAG GCCGGCGAGC AGATTTCCCA 420
AGGCGCTGAT CACGTGCTGT GTGTCGCGGC AGAGTCCATG TCGCGTAACC CCATCGCGTC 480
GTATACACAC CGGGGCGGGT TCCGCCTCGG TGCGCCCGTT GAGTTCAAGG ATTTTTTGTG 540
GGAGGCATTG TTTGATCCTG CTCCAGGACT CGACATGATC GCTACCGCAG AAAACCTGGG 600
GGAGAGGCGG TTTGCGTATT GGGCGCATGC ATAAAAACTG TTGTAATTCA TTAAGCATTC 660
TGCCGACATG GAAGCCATCA CAAACGGCAT GATGAACCTG AATCGCCAGC GGCATCAGCA 720
CCTTGTCGCC TTGCGTATAA TATTTGCCCA TGGACGCACA CCGTGGAAAC GGATGAAGGC 780
ACGAACCCAG TTGACATAAG CCTGTTCGGT TCGTAAACTG TAATGCAAGT AGCGTATGCG 840
CTCACGCAAC TGGTCCAGAA CCTTGACCGA ACGCAGCGGT GGTAACGGCG CAGTGGCGGT 900
TTTCATGGCT TGTTATGACT GTTTTTTTGT ACAGTCTATG CCTCGGGCAT CCAAGCAGCA 960
AGCGCGTTAC GCCGTGGGTC GATGTTTGAT GTTATGGAGC AGCAACGATG TTACGCAGCA 1020
GCAACGATGT TACGCAGCAG GGCAGTCGCC CTAAAACAAA GTTAGGTGGC TCAAGTATGG 1080
GCATCATTCG CACATGTAGG CTCGGCCCTG ACCAAGTCAA ATCCATGCGG GCTGCTCTTG 1140
ATCTTTTCGG TCGTGAGTTC GGAGACGTAG CCACCTACTC CCAACATCAG CCGGACTCCG 1200
ATTACCTCGG GAACTTGCTC CGTAGTAAGA CATTCATCGC GCTTGCTGCC TTCGACCAAG 1260
AAGCGGTTGT TGGCGCTCTC GCGGCTTACG TTCTGCCCAG GTTTGAGCAG CCGCGTAGTG 1320
AGATCTATAT CTATGATCTC GCAGTCTCCG GCGAGCACCG GAGGCAGGGC ATTGCCACCG 1380
CGCTCATCAA TCTCCTCAAG CATGAGGCCA ACGCGCTTGG TGCTTATGTG ATCTACGTGC 1440
AAGCAGATTA CGGTGACGAT CCCGCAGTGG CTCTCTATAC AAAGTTGGGC ATACGGGAAG 1500
AAGTGATGCA CTTTGATATC GACCCAAGTA CCGCCACCTA ACAATTCGTT CAAGCCGAGA 1560
TCGGCTTCCC ATTGAGGGCG CAAGAGGAGA AATGGATTGA CCAAGAGATC GTGGCTGTTA 1620
CGGATGAACA GTTCGATTTA GAGGGCTACA ACAGTCGAGC AATTGAACTG CCTCGGAAGG 1680
CAAAATTGTT GATCGTGACA GTCATCCGCG GCCTAGCAGT CTTTGAAGCC CTTTCCCGAT 1740
TGAAGCCTGT TCATTCTGGC GGGGTGCAGA CTGCGGGCAA CAGCTGTGCC GTAGTGGACG 1800
GCGCCGCGGC GGCTTTGGTG GCTCGAGAGT CGTCTGCGAC ACAGCCGGTC TTGGCTAGGA 1860
TACTGGCTAC CTCCGTAGTC GGGATCGAGC CCGAGCATAT GGGGCTCGGC CCTGCGCCCG 1920
CGATTCGCCT GCTGCTTGCG CGTAGTGATC TTAGTTTGAG GGATATCGAC CTCTTTGAGA 1980
TAAACGAGGC GCAGGCCGCC CAAGTTCTAG CGGTACAGCA TGAATTGGGT ATTGAGCACT 2040
CAAAACTTAA TATTTGGGGC GGGGCCATTG CACTTGGACA CCCGCTTGCC GCGACCGGAT 2100
TGCGTCTCTG CATGACCCTC GCTCACCAAT TGCAAGCTAA TAACTTTCGA TATGGAATTG 2160

```
CCTCGGCATG CATTGGTGGG GGACAGGGGA TGGCGGTTCT TTTAGAGAAT CCCCACTTCG  2220
GTTCGTCCTC TGCACGAAGT TCGATGATTA ACAGAGTTGA CCACTATCCA CTGAGCTAAC  2280
GGGCATCTCC TTTGTTGCTT TGAGGTGGCG CACGAAGGAG GGCTCGAAAA TCTCTGCTAA  2340
AAACAAGAAG AAGGAACAGG GAACATGATT AGTTTCGCTC GTATGGCAGA AAGTTTAGGA  2400
GTCCAGGCTA AACTTGCCCT TGCCTTCGCA CTCGTATTAT GTGTCGGGCT GATTGTTACC  2460
GGCACGGGTT TCTACAGTGT ACATACCTTG TCAGGGTTGG TGGGAATTC              2509
```

Sequence 18

```
GAATTCCGCG GTCGGCGAAA GTTGATGCGC TGTATCGTGG TGAAGATCAA TCCATGCTGC     60
GTGACGAGGC CACACTGTGA GTTGGTCAGG GGGGGCTTAC TCGGCGTTTT CCGACACTGC    120
GTTGGTTGCG GCAGTGCGCA CCCCCTGGAT TGATTGCGGG GGTGCCCTGT CGCTGGTGTC    180
GCCTATCGAC TTAGGGGTAA AGGTCGCTCG CGAAGTTCTG ATGCGTGCGT CGCTTGAACC    240
ACAAATGGTC GATAGCGTAC TCGCAGGCTC TATGGCTCAA GCAAGCTTTG ATGCTTACCT    300
GCTCCCGCGG CACATTGGCT TGTACAGCGG TGTTCCCAAG TCGGTTCCGG CCTTGGGGGT    360
GCAGCGCATT TGCGGCACAG GCTTCGAACT GCTTCGGCAG GCCGGCGAGC AGATTTCCCA    420
AGGCGCTGAT CACGTGCTGT GTGTCGCGGC AGAGTCCATG TCGCGTAACC CCATCGCGTC    480
GTATACACAC CGGGGCGGGT TCCGCCTCGG TGCGCCCGTT GAGTTCAAGG ATTTTTTGTG    540
GGAGGCATTG TTTGATCCTG CTCCAGGACT CGACATGATC GCTACCGCAG AAAACCTGGC    600
GCGCATTGAG GGCGCAAGAG GAGAAATGGA TTGACCAAGA GATCGTGGCT GTTACGGATG    660
AACAGTTCGA TTTAGAGGGC TACAACAGTC GAGCAATTGA ACTGCCTCGG AAGGCAAAAT    720
TGTTGATCGT GACAGTCATC CGCGGCCTAG CAGTCTTTGA AGCCCTTTCC CGATTGAAGC    780
CTGTTCATTC TGGCGGGGTG CAGACTGCGG GCAACAGCTG TGCCGTAGTG GACGGCGCCG    840
CGGCGGCTTT GGTGGCTCGA GAGTCGTCTG CGACACAGCC GGTCTTGGCT AGGATACTGG    900
CTACCTCCGT AGTCGGGATC GAGCCCGAGC ATATGGGGCT CGGCCCTGCG CCCGCGATTC    960
GCCTGCTGCT TGCGCGTAGT GATCTTAGTT TGAGGGATAT CGACCTCTTT GAGATAAACG   1020
AGGCGCAGGC CGCCCAAGTT CTAGCGGTAC AGCATGAATT GGGTATTGAG CACTCAAAAC   1080
TTAATATTTG GGGCGGGGCC ATTGCACTTG GACACCCGCT TGCCGCGACC GGATTGCGTC   1140
TCTGCATGAC CCTCGCTCAC CAATTGCAAG CTAATAACTT TCGATATGGA ATTGCCTCGG   1200
CATGCATTGG TGGGGGACAG GGGATGGCGG TTCTTTTAGA GAATCCCCAC TTCGGTTCGT   1260
CCTCTGCACG AAGTTCGATG ATTAACAGAG TTGACCACTA TCCACTGAGC TAACGGGCAT   1320
CTCCTTTGTT GCTTTGAGGT GGCGCACGAA GGAGGGCTCG AAAATCTCTG CTAAAAACAA   1380
GAAGAAGGAA CAGGGAACAT GATTAGTTTC GCTCGTATGG CAGAAAGTTT AGGAGTCCAG   1440
GCTAAACTTG CCCTTGCCTT CGCACTCGTA TTATGTGTCG GGCTGATTGT TACCGGCACG   1500
GGTTTCTACA GTGTACATAC CTTGTCAGGG TTGGTGGGAA TTC                     1543
```

Fig. 1a

Fig. 1b

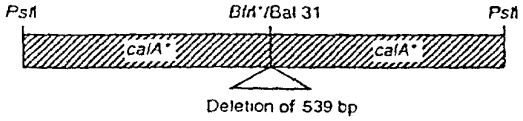


Fig. 1c

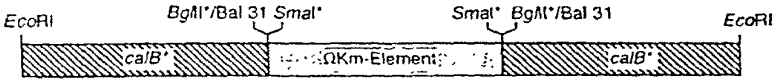


Fig. 1d

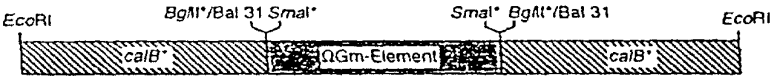


Fig. 1e

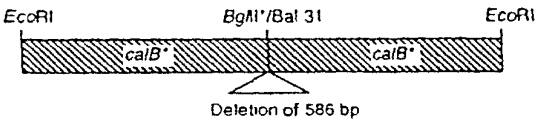


Fig. 1f

*fcs*ΩKm

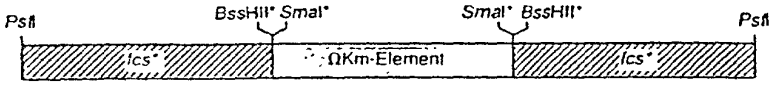


Fig. 1g

*fcs*ΩGm

Fig. 1h

*fcs*Δ

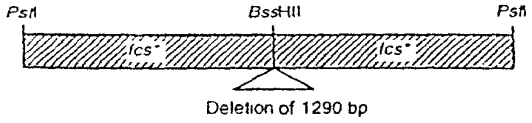


Fig. 1i

*ech*ΩKm

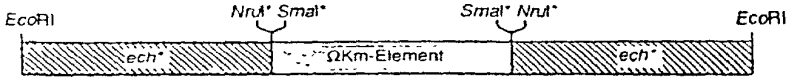


Fig. 1j

*ech*ΩGm

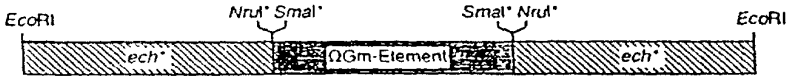


Fig. 1k

*ech*Δ

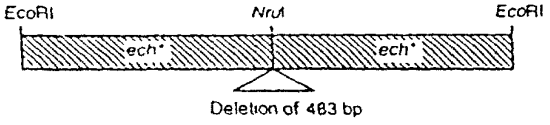
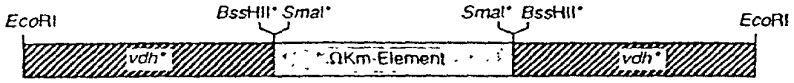
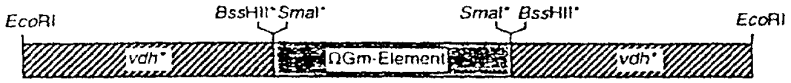


Fig. 1l



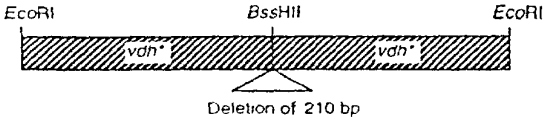
*vdh*ΩKm

Fig. 1m



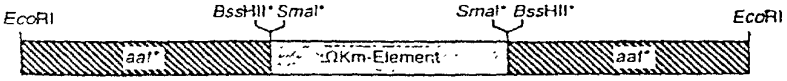
*vdh*ΩGm

Fig. 1n



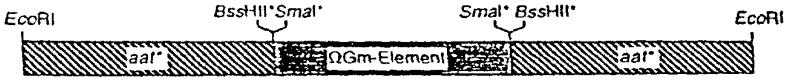
*vdh*Δ

Fig. 1o



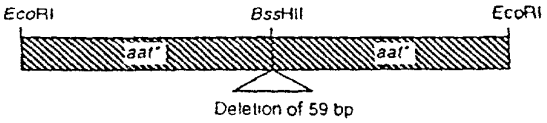
*aat*ΩKm

Fig. 1p



*aat*ΩGm

Fig. 1q



*aat*Δ

Fig. 1r

COMBINED DECLARATION AND POWER OF ATTORNEY	ATTORNEY DOCKET NO

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name. I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought
on the invention entitled

"CONSTRUCTION OF PRODUCTION STRAINS FOR PRODUCING SUBSTITUTED PHENOLS BY SPECIFICALLY INACTIVATING GENES OF THE EUGENOL AND FERULIC ACID CATABOLISM"

the specification of which is attached hereto,

or was filed on **October 20, 1999**

as a PCT Application Serial No. **PCT/EP99/07952**

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims.

I acknowledge the duty to disclose information which is material to the patentability of this application in accordance with Title 37, Code of Federal Regulations, §1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

Prior Foreign Application(s), the priority(ies) of which is/are to be claimed:

198 50 242.7	**Germany**	**October 31, 1998**
(Number)	(Country)	(Month/Day/Year Filed)

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose the material information as defined in Title 37, Code of Federal Regulations, §1.56 which occured between the filing date of the prior application and the national or PCT international filing date of this application:

(Application Serial No.)	(Filing Date)	(Status) (patented, pending, abandoned)
(Application Serial No.)	(Filing Date)	(Status) (patented, pending, abandoned)

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith:

JOSEPH C. GIL, Patent Office Registration Number 26,602
ARON PREIS, Patent Office Registration Number 29,426
LYNDANNE M. WHALEN, Patent Office Registration Number 29,457
THOMAS W. ROY, Patent Office Registration Number 29,582
RICHARD E. L. HENDERSON, Patent Office Registration Number 31,619
GODFRIED R. AKORLI, Patent Office Registration Number 28,779
N. DENISE BROWN, Patent Office Registration Number 36,097
NOLAND J. CHEUNG, Patent Office Registration Number 39,138
DIDERICO VAN EYL, Patent Office Registration Number 38,641
CAROLYN M. SLOANE, Patent Office Registration Number 44,339
JAMES R. FRANKS, Patent Office Registration Number 42,552
JACKIE ANN ZURCHER, Patent Office Registration Number 42,251
all of Bayer Corporation, Pittsburgh, Pennsylvania 15205-9741

Send Correspondence To:	**Direct Telephone Calls To:**
Patent Department **Bayer Corporation** **100 Bayer Road** **Pittsburgh, Pennsylvania 15205-9741**	**(412) 777-2349**

FULL NAME OF SOLE OR FIRST INVENTOR	INVENTOR'S SIGNATURE	DATE
Jürgen Rabenhorst	Jürgen Rabenhorst	30.3.2001
RESIDENCE		CITIZENSHIP
D 37671 Höxter, Germany		German
POST OFFICE ADDRESS		
c/o HAARMANN & REIMER GMBH, D 37601 Holzminden, Germany		
FULL NAME OF SECOND INVENTOR	INVENTOR'S SIGNATURE	DATE
Alexander Steinbüchel		17.4.2001
RESIDENCE		CITIZENSHIP
D 48341 Altenberge, Germany		German
POST OFFICE ADDRESS		
Rönnenthal 27, D 48341 Altenberge, Germany		
FULL NAME OF THIRD INVENTOR	INVENTOR'S SIGNATURE	DATE
Horst Priefert		19.4.2001
RESIDENCE		CITIZENSHIP
D 48291 Telgte, Germany		German
POST OFFICE ADDRESS		
Potthoffskamp 5, D 48291 Telgte, Germany		
FULL NAME OF FOURTH INVENTOR	INVENTOR'S SIGNATURE	DATE
Jörg Overhage		18.4.2001
RESIDENCE		CITIZENSHIP
D 59192 Bergkamen, Germany		German
POST OFFICE ADDRESS		
Rotherbachstr. 42A, D 59192 Bergkamen, Germany		
FULL NAME OF FIFTH INVENTOR	INVENTOR'S SIGNATURE	DATE
RESIDENCE		CITIZENSHIP
POST OFFICE ADDRESS		
FULL NAME OF SIXTH INVENTOR	INVENTOR'S SIGNATURE	DATE
RESIDENCE		CITIZENSHIP
POST OFFICE ADDRESS		
FULL NAME OF SEVENTH INVENTOR	INVENTOR'S SIGNATURE	DATE
RESIDENCE		CITIZENSHIP
POST OFFICE ADDRESS		